

高职高专计算机规划教材

多媒体技术教程

郑阿奇 主 编

刘 毅 编 著

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书主要介绍多媒体计算机系统、数字声音、数字图像、数字视频、多媒体数据压缩、数字动画、多媒体电子出版物与创作工具和数字流媒体等。全书包含两部分：实用教程和操作实验。实用教程理论联系实际，为实际应用服务，操作实验突出实际操作应用。“学用结合”是本书介绍多媒体的主要目标，以基本理论为主，介绍主要内容时都尽可能地跟实际应用相结合。

本书可作为高职高专有关课程的教材，也可作为广大多媒体应用人员的参考书。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

图书在版编目(CIP)数据

多媒体技术教程 / 郑阿奇主编. —北京：电子工业出版社，2010.1

（高职高专计算机规划教材）

ISBN 978-7-121-09682-2

I.多... II.郑... III.多媒体技术—高等学校：技术学校—教材 IV.TP37

中国版本图书馆 CIP 数据核字（2009）第 184730 号

策划编辑：赵云峰

责任编辑：徐云鹏

印 刷：

装 订：

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本：787×1092 1/16 印张：20 字数：512 千字

印 次：2010 年 1 月第 1 次印刷

印 数：4 000 册 定价：29.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zltz@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

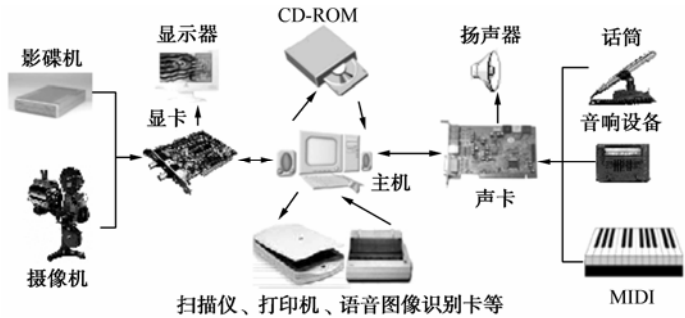
前言

多媒体是高职高专许多专业必须开设的课程，为此，我们首先对目前市场上高职高专的多媒体教材进行了深入分析，找出其主要特点和存在的问题，对如何让学生“学用结合”，如何方便教师教学进行了专门研究，在继承实用教程系列的成功经验基础上，针对高职高专的教学情况，编写了此书。

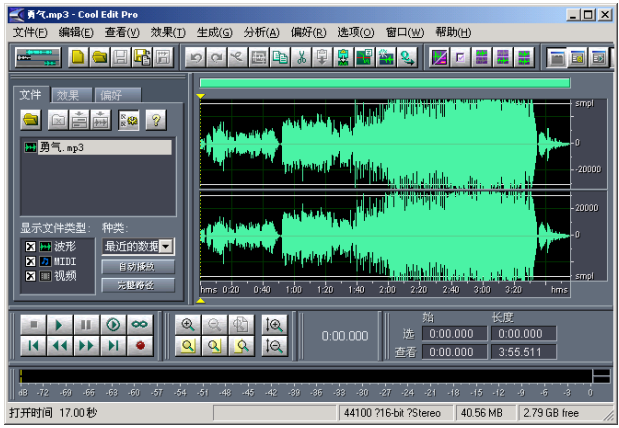
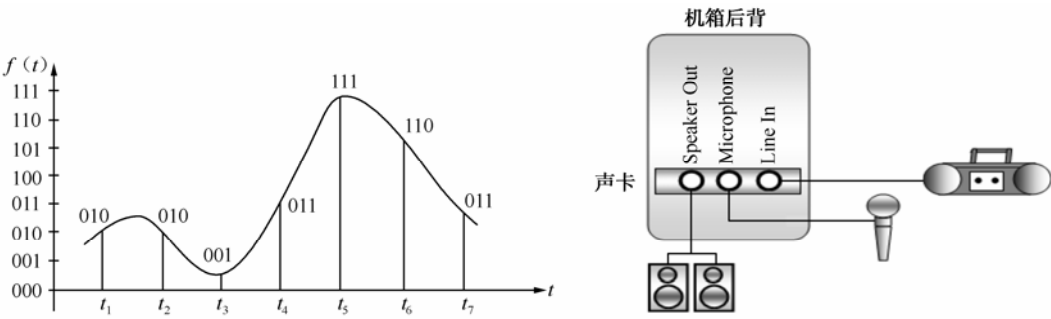
本书第一部分：实用教程，理论联系实际，为实际应用服务，并且通过实例的形式表现出来。第二部分：操作实验，实际操作应用，对部分知识消理解。

本书介绍的多媒体的主要内容和风格如下：

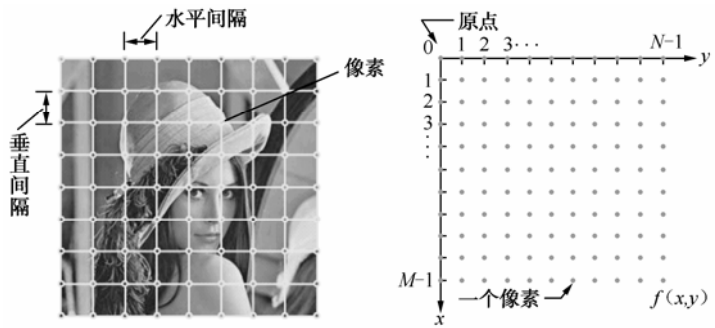
(1) 多媒体计算机



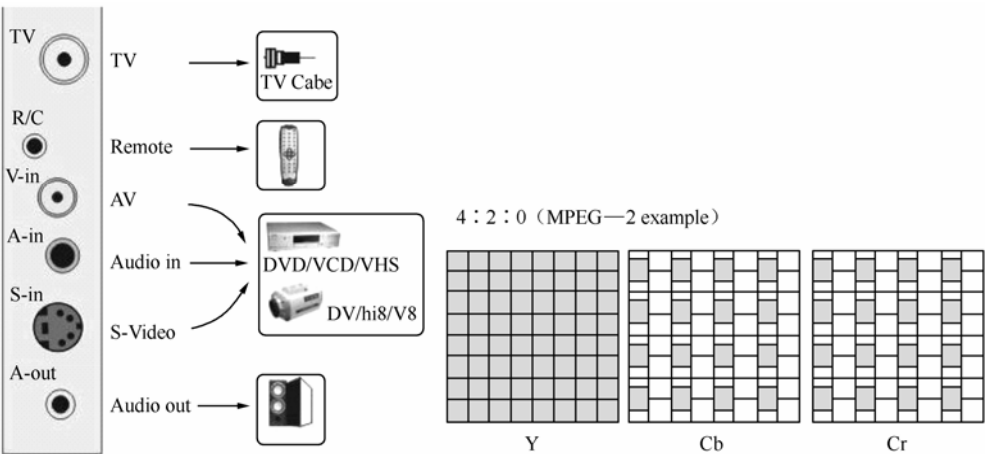
(2) 数字声音处理



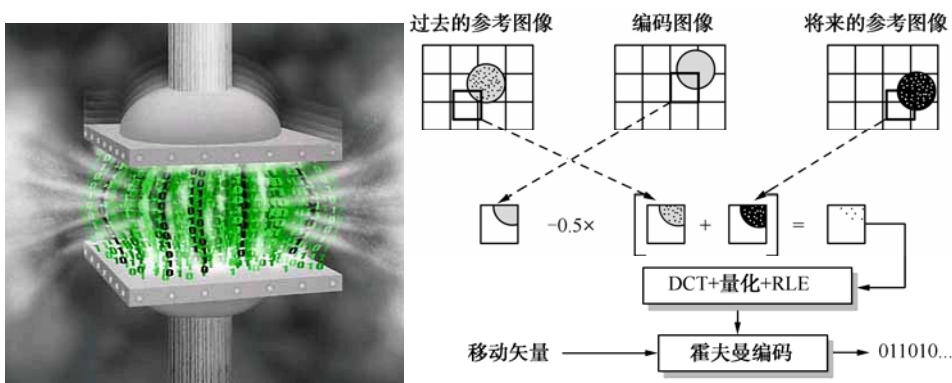
(3) 数字图像处理



(4) 数字视频处理



(5) 多媒体数据压缩



本教程由电子工业出版社的教学服务平台 (<http://www.hxedu.com.cn>) 为读者提供服务, 可免费下载教学课件、实例源文件等资料。

本书由刘毅编写, 郑阿奇对全书进行统编、定稿。其他很多同志也对本书的编写提供了许多帮助, 在此一并表示感谢!

由于作者水平有限, 不当之处在所难免, 恳请读者批评指正。

目 录

第一部分 实 用 教 程

第 1 章 多媒体技术概论	(1)
1.1 多媒体的基本概念	(1)
1.1.1 媒体的分类	(1)
1.1.2 多媒体定义	(2)
1.2 多媒体计算机系统	(4)
1.2.1 多媒体计算机硬件系统	(4)
1.2.2 多媒体计算机软件系统	(5)
1.3 多媒体关键技术	(6)
1.4 多媒体技术的应用及其发展趋势	(9)
1.4.1 多媒体技术的应用	(9)
1.4.2 多媒体技术的发展趋势	(10)
习题 1	(11)
第 2 章 数字声音	(12)
2.1 声音信号数字化	(12)
2.1.1 声音基础	(12)
2.1.2 声音信号数字化过程	(13)
2.1.3 声音数字化的主要参数	(15)
2.1.4 数字化声音的压缩	(18)
2.1.5 声音文件和格式	(18)
2.2 声卡	(21)
2.2.1 声卡介绍	(21)
2.2.2 声卡的工作原理	(24)
2.3 声音工具软件	(27)
2.4 MIDI	(28)
2.4.1 MIDI 规范与音乐合成	(29)
2.4.2 MIDI 系统	(32)
2.4.3 MIDI 消息、文件和格式	(35)
2.4.4 MIDI 工具软件	(36)
2.5 其他语音技术	(37)
2.5.1 语音合成	(38)
2.5.2 语音识别技术	(39)
2.5.3 开发语音识别或语音合成方面的应用程序	(41)
2.5.4 CD 的数据	(41)
习题 2	(42)
第 3 章 数字图像	(44)

3.1	认识色彩	(44)
3.1.1	色彩的来源	(44)
3.1.2	色彩的三要素	(45)
3.1.3	色彩模型	(46)
3.2	图像数据的获取与表示	(50)
3.3	图像的基本属性	(52)
3.3.1	图像的分辨率	(52)
3.3.2	图像的像素深度	(54)
3.3.3	图像的大小	(55)
3.3.4	图像的真彩色、伪彩色与直接色	(55)
3.3.5	图像的种类	(56)
3.4	图像的文件格式	(58)
3.4.1	BMP 格式	(58)
3.4.2	GIF 格式	(61)
3.4.3	JPEG 格式	(64)
3.4.4	TIFF 格式	(66)
3.4.5	PSD 格式	(66)
3.5	图像系统组成	(66)
3.5.1	显示器	(66)
3.5.2	扫描仪	(76)
3.5.3	数码相机	(78)
3.5.4	打印机	(80)
3.6	数字图像处理	(84)
3.6.1	数字图像处理的内容	(84)
3.6.2	数字图像处理的应用	(89)
3.7	图像处理工具软件	(90)
	习题 3	(93)
第 4 章	数字视频	(95)
4.1	视频基本概念	(95)
4.1.1	模拟视频	(95)
4.1.2	数字视频	(101)
4.2	视频的数字化过程	(101)
4.2.1	视频信号的采样	(101)
4.2.2	视频信号的量化	(105)
4.2.3	视频信号的压缩与编码	(105)
4.3	数字视频的获取	(105)
4.3.1	数字视频系统的组成	(106)
4.3.2	视频采集卡的工作原理	(106)
4.3.3	视频采集卡的性能指标	(107)
4.3.4	视频采集过程	(109)
4.3.5	其他功能的视频设备	(111)

4.4	数字视频的编辑	(115)
4.4.1	视频编辑术语	(115)
4.4.2	视频编辑软件的基本功能	(116)
4.5	数字视频的文件格式	(117)
4.5.1	本地视频格式	(117)
4.5.2	网络(流式)视频格式	(119)
4.6	数字视频文件格式转换	(121)
4.6.1	全能转换工具	(121)
4.6.2	RM 相关转换工具	(124)
4.6.3	DivX 相关转换工具	(125)
4.6.4	DVD 相关转换工具	(126)
	习题 4	(126)
第 5 章	多媒体数据压缩	(128)
5.1	数据压缩的基本概念	(128)
5.1.1	为什么要数据压缩	(128)
5.1.2	多媒体数据压缩的发展	(129)
5.1.3	多媒体数据压缩的分类	(130)
5.1.4	多媒体数据压缩性能的评价标准	(130)
5.2	静态图像压缩标准 JPEG	(131)
5.2.1	JPEG 算法概要	(132)
5.2.2	JPEG 算法的主要步骤	(132)
5.2.3	JPEG2000 简介	(138)
5.3	运动图像压缩标准 MPEG	(139)
5.3.1	MPEG—1	(140)
5.3.2	MPEG—2	(144)
5.3.3	MPEG—4	(148)
5.3.4	MPEG—7	(151)
5.3.5	MPEG—21	(153)
5.4	H.26x	(155)
5.4.1	H.261	(155)
5.4.2	H.262	(157)
5.4.3	H.263	(157)
5.4.4	H.264	(157)
	习题 5	(158)
第 6 章	数字动画	(160)
6.1	动画基本概念	(160)
6.1.1	动画的视觉原理	(160)
6.1.2	计算机动画的概念	(160)
6.1.3	计算机动画的分类	(161)
6.1.4	计算机动画的制作	(164)

6.1.5	计算机动画的应用	(166)
6.1.6	计算机动画的发展趋势	(170)
6.2	GIF 动画	(170)
6.3	Flash 动画	(171)
6.3.1	Flash 动画的制作方式	(171)
6.3.2	Flash 动画的制作软件	(172)
6.4	三维动画制作	(172)
6.4.1	三维动画的原理	(172)
6.4.2	三维动画的制作软件	(173)
习题 6		(175)
第 7 章	多媒体电子出版物与创作工具	(176)
7.1	多媒体电子出版物	(176)
7.1.1	多媒体电子出版物的概念与特点	(176)
7.1.2	多媒体电子出版物的开发与制作	(177)
7.2	多媒体创作工具	(177)
7.2.1	多媒体创作工具的特点	(177)
7.2.2	多媒体创作工具的分类	(178)
7.3	常用多媒体创作工具介绍	(180)
习题 7		(181)
第 8 章	数字流媒体	(182)
8.1	流媒体基本概念	(182)
8.1.1	流媒体技术概述	(182)
8.1.2	流媒体文件格式	(187)
8.1.3	流媒体技术的应用	(190)
8.2	流媒体处理技术	(192)
8.2.1	Real System	(192)
8.2.2	Windows Media Services	(194)
8.2.3	QuickTime	(197)
8.3	流媒体的下载	(198)
8.3.1	HiDownload	(198)
8.3.2	Net Transport	(199)
习题 8		(200)
第 9 章	多媒体光盘存储系统	(202)
9.1	光盘存储系统	(202)
9.1.1	光盘及其特点	(202)
9.1.2	光盘的类型	(203)
9.1.3	光盘驱动器	(204)
9.2	光盘的标准	(206)
9.2.1	CD-DA 标准 (红皮书标准)	(207)
9.2.2	CD-ROM 标准 (黄皮书标准)	(207)

9.2.3	CD-R 标准（橙皮书标准）	(207)
9.2.4	Video-CD 标准（白皮书标准）	(208)
9.2.5	DVD 标准	(210)
习题 9		(217)
第 10 章	网络多媒体技术	(219)
10.1	超文本与超媒体	(219)
10.1.1	超文本/超媒体的基本概念	(219)
10.1.2	超文本/超媒体组成要素	(219)
10.1.3	超文本/超媒体的实现机制——HTML 语言	(220)
10.2	多媒体通信网络	(227)
10.2.1	多媒体通信的发展背景	(227)
10.2.2	多媒体通信的特点	(228)
10.2.3	多媒体通信的关键技术	(229)
10.3	网络多媒体应用	(229)
10.3.1	网络多媒体应用概述	(229)
10.3.2	视频会议系统	(231)
10.3.3	视频点播系统	(233)
10.3.4	IP 电话	(235)
习题 10		(238)

第二部分 操作实验

实验 1	数字声音	(240)
实验目的		(240)
实验内容		(240)
T1.1	Windows 录音机获取声音实例	(240)
T1.2	Easy CD-DA Extractor 获取 CD 声音实例	(243)
T1.3	CoolEdit Pro 制作个人音乐实例	(245)
实验思考		(254)
实验 2	数字图像	(255)
实验目的		(255)
实验内容		(255)
T2.1	Windows 屏幕图像获取、加工和使用实例	(255)
T2.2	Photoshop 图像处理实例	(258)
实验思考		(271)
实验 3	数字视频	(272)
实验目的		(272)
实验内容		(272)
T3.1	视频的拍摄与 Premiere 的使用	(272)
实验思考		(278)
实验 4	数字动画	(279)
实验目的		(279)

实验内容..... (279)

 T4.1 Ulead GIF Animator 制作 GIF 动画实例..... (279)

 T4.2 Flash 制作动画实例..... (287)

实验思考..... (298)

实验 5 多媒体创作工具..... (299)

 实验目的..... (299)

 实验内容..... (299)

 T5.1 Authorware 简介 (299)

 T5.2 Authorware 多媒体创作实例..... (302)

 实验思考..... (308)

第一部分 实用教程

第 1 章 多媒体技术概论

20 世纪 80 年代，随着微电子、计算机和数字化声像技术的飞速发展，多媒体技术应运而生。它的应用使计算机人-机界面集声音、文字、图像于一体，使用户置身于多种媒体协同工作的环境中，对人类的生产和生活方式带来了巨大的变革，使人类社会进入到一个前所未有的新时代。

1.1 多媒体的基本概念

1.1.1 媒体的分类

媒体（Medium）在计算机领域中有两种含义，一种是指用以存储信息的实体，如磁带、磁盘、光盘和半导体存储器；另一种是指信息的载体，如数字、文字、声音、图形和图像。多媒体技术中的媒体是指后者。

国际电话电报咨询委员会 CCITT（Consultative Committee on International Telephone and Telegraph，国际电信联盟 ITU 的一个分会）把媒体分成如下 5 类。

（1）感觉媒体（Perception Medium）

感觉媒体就是指能直接作用于人的感觉器官，使人直接产生感觉（视、听、嗅、味、触觉）的一类媒体，如语言、音乐、图形、动画、数据、文字、文件等都是感觉媒体。

（2）表示媒体（Representation Medium）

表示媒体是为了更有效地加工、处理和传输感觉媒体而人为研究和构造出来的一种媒体。它包括上述感觉媒体的各种编码，诸如语言编码、静止和活动图像编码，以及文本编码等。

（3）表现媒体（Presentation Medium）

表现媒体是感觉媒体与用于通信的电信号之间在转换中用于信息输入和输出的媒体。可分为输入表现媒体（如键盘、摄像机、话筒、扫描仪等）和输出表现媒体（如显示器、扬声器、打印机等）两种。

（4）存储媒体（Storage Medium）

存储媒体用于存储表示媒体，即存放感觉媒体数字化后的代码的媒体称为存储媒体，如磁盘、光盘、磁带、半导体存储器等。

（5）传输媒体（Transmission Medium）

传输媒体是指用来将表示媒体从一处传递到另一处的物理传输介质，如同轴电缆、双绞线、光纤及其他通信信道。

1.1.2 多媒体定义

1. 多媒体

关于多媒体（Multimedia）的定义或说法多种多样，从不同的角度出发对多媒体给出了不同的描述，目前仍没有统一的标准。通常所说的多媒体就是各种媒体的综合，也就是声音、图像、动画、文字、数据、文件等各种媒体的综合。

从广义上来讲，多媒体是指多种信息媒体的表现和传播形式。人们在日常生活中进行交流时，可以以声音、文字、图像、手势和体态来进行信息传递，还可以通过嗅觉、味觉、视觉和触觉系统来感受外界信息，因此从某种意义上讲，人是一个多媒体信息处理系统。

从狭义上来讲，多媒体是指人们用计算机及其他设备交互处理多媒体信息的方法和手段，或指在计算机中处理多种媒体的一系列技术。

2. 多媒体技术

多媒体技术是一种基于计算机科学的综合技术，它包括数字化信息处理技术、音频和视频技术、计算机软件和硬件技术、人工智能和模式识别技术、通信和网络技术等。简单地说，所谓多媒体技术是指计算机交互式综合处理多媒体信息——文本、图形、图像和声音等，使多种信息建立逻辑连接，集成为一个系统并具有交互性的一项技术，如图 1.1 所示。



图 1.1 多媒体技术示意图

多媒体技术强调的是交互式综合处理多种媒体（尤其是感觉媒体）的技术。从本质上来看，它具有信息载体的多样性、集成性和交互性这 3 个主要特征。

（1）多样性

多样性是相对于计算机而言的，指信息媒体的多样性，又称多维化。信息媒体多样性使计算机所能处理的信息范围从传统的数值、文字、静止图像扩展到声音和视频信息。

（2）集成性

集成性又称综合性。多媒体技术的集成性主要表现在两个方面，即多种信息媒体的集成和处理这些媒体的设备的集成。对于前者而言，这种集成包括信息的多通道统一获取、多媒体信息的统一存储与组织、多媒体信息表现与合成等各个方面，而不像早期那样，只使用单一的形式。多媒体的各种设备也应该集成为一体，从硬件上来说，应该具有能够处理多媒体信息的高性能计算机系统和与之相对应的输入/输出能力及外设；从软件上来说，应该有集成一体化的多媒体操作系统，适合于多媒体信息管理和使用的系统软件、创作工具及各类应用软件等。总之，集成性能使多种不同形式的信息综合地表现某个内容，从而取得更好的效果。

（3）交互性

交互性是多媒体技术的关键特征，它使人们获取和使用信息时变被动为主动，可以更有效地控制和使用信息，增加对信息的注意和理解。众所周知，一般的电视机是声像一体化的，是把多种媒体集成在一起的设备，但它不具备交互性，因为用户只能使用信息，而不能自由地控制和处理信息。当引入多媒体技术后，借助交互性，用户可以获得更多的信息。例如，

在多媒体远程信息检索系统中，初级交互性可使用户找出想读的书籍，快速跳过不感兴趣的部分，从数据库中检索声音、图像或文字材料等。中级交互性则可使用户介入到信息的提取和处理过程中，如对关心的内容进行编排、插入文字说明及解说等。当采用虚拟现实（Virtual Reality）技术时，多媒体系统可提供高级的交互性。

3. 超媒体

超媒体（Hypermedia）起源于超文本（Hypertext）。超文本可以简单地定义为收集、存储和浏览离散信息，以及建立和表示信息之间的关系的技術，它与传统的线性文本结构有很大差别，它符合人类的“联想”式思维习惯。

超文本的概念可用图 1.2 来说明。超文本中带有链接关系的文本通常用下画线和不同的颜色表示。文本①中的“Web”与②中的“Web”建立有链接关系，①中的“因特网”与③中的“因特网”建立有链接关系……，建立互相链接的这些对象不受空间位置的限制，它们可以在同一个文件内也可以在不同的文件之间，还可以通过网络与世界上的任何一台连网计算机上的文件建立链接关系。

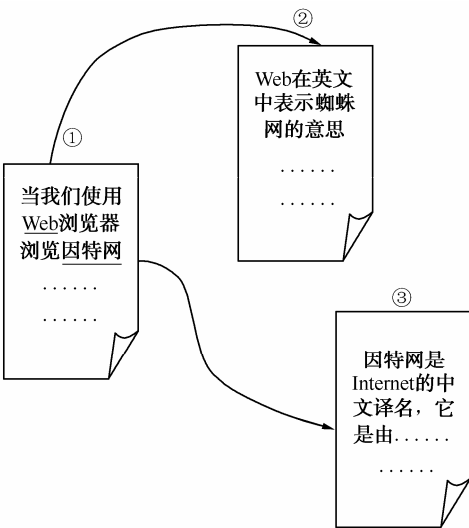


图 1.2 超文本的概念

超文本可引用、链接其他不同类型（内含声音、图片、动画等）的文件，这些具有多媒体操作的超文本，称为超媒体，意指多媒体超文本（Multimedia Hypertext），即以多媒体的方式呈现相关文件信息。超文本和超媒体只是研究对象不同，所以一般并不区别它们。

超文本/超媒体技术的出现，为实现多媒体信息综合有效的管理带来了希望，尤其在 Internet 飞速发展的今天，超文本/超媒体技术已成为 Internet 上信息检索的核心技术。

4. 流媒体

流媒体（Streaming Media）其实是一种多媒体文件，其在网络上传输的过程中应用了流技术。所谓流技术，就是把完整的影像和声音数据经过压缩处理后保存在网站服务器上，用户可以边下载边获取信息，从而无须将整个压缩文件下载之后再观看的网络传输技术。与单纯的下载方式相比，这种对多媒体文件边下载边播放的流式传输方式不仅使启动延时大幅度

缩短，而且对系统缓存容量的需求也大大降低。

1.2 多媒体计算机系统

多媒体计算机系统是一个能综合处理多种媒体信息的系统，由多媒体计算机硬件系统和多媒体计算机软件系统构成。多媒体计算机硬件系统的核心是一台高性能的计算机系统，外部设备主要由音频、视频和存储设备组成。多媒体计算机软件系统主要包括多媒体操作系统、多媒体创作工具软件和多媒体应用系统。

1.2.1 多媒体计算机硬件系统

多媒体硬件是多媒体计算机系统的基本物质实体，其主要任务是能够实时地综合处理文、图、声、像信息，实现全动态视像和立体声的处理，同时还需对多媒体信息进行实时压缩与解压缩。

1. 多媒体计算机硬件系统的基本组成

构成多媒体硬件系统除了需要较高配置的计算机主机硬件以外，通常还需要音频/视频处理设备、光盘驱动器、各种媒体输入/输出设备等。一个具有基本功能的多媒体计算机硬件系统如图 1.3 所示。

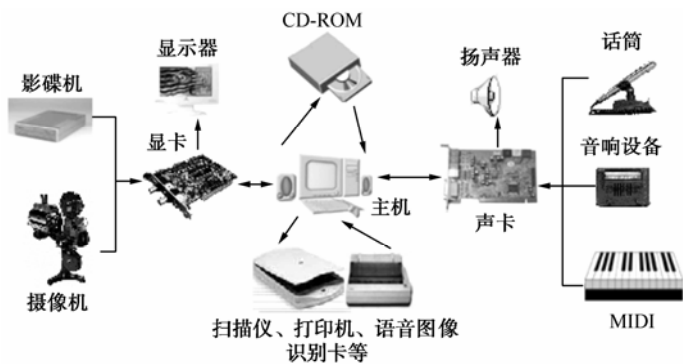


图 1.3 一个具有基本功能的多媒体计算机硬件系统

(1) 主机

多媒体计算机主机可以是中、大型机，也可以是工作站，然而目前更普遍的是多媒体个人计算机，即 MPC (Multimedia Personal Computer)。

(2) 多媒体接口卡

多媒体接口卡是根据多媒体系统获取、编辑音频或视频的需要插接在计算机上，以解决各种媒体数据的输入/输出问题的。常用的接口卡有声卡、显卡、视频压缩卡、视频解压卡、视频捕捉卡等。

(3) 多媒体外部设备

多媒体外部设备的工作方式一般为输入和输出两类。按其功能又可分如下 4 类。

- 视频、音频输入设备：如摄像机、录像机、扫描仪、传真机、数字相机、话筒等。
- 视频、音频输出设备：如显示器、投影仪、音响等。

- 人-机交互设备：如键盘、鼠标、触摸屏、绘图板、光笔及手写输入设备等。
- 存储设备：如磁盘、光盘等。

2. 多媒体计算机的硬件标准

多媒体技术是一项综合技术，多媒体产品统一标准的制定将推动相关产业的快速发展，产品的成本与价格大幅度下降，产品的兼容性和通用性大大改善，反过来又促进应用的迅速发展。

为促进多媒体个人计算机 MPC (Multimedia Personal Computer) 的标准化，1990 年 Microsoft 公司联合几个主要个人计算机厂商组成了 MPC 市场协会 (Multimedia PC Marketing Council)，并在 1991 年推出了第一代多媒体计算机 MPC 的标准，在 1993 年、1995 年分别推出了 MPC 2.0，MPC 3.0。随着技术的不断发展，1996 年又发表了 MPC 4.0 的技术标准及后来的 MPC 5.0 等。

按照 MPC 标准，多媒体计算机包含 5 个基本单元：个人计算机、CD-ROM 驱动器、声卡、Windows 操作系统及一组音箱或耳机。同时它对个人计算机的 CPU、内存、硬盘、显示功能等提出了基本要求。个人计算机只要符合这些规定就可获得 MPC 认证，并可使用 MPC 标志。

MPC 5.0 主要技术规范包括 CPU——Pentium II/III，内存——32MB/64MB，硬盘——3.1GB/6.4GB，CD-ROM——20x/40x，声卡——16/32 位，分辨率——1280×1024 和图像——32 位真彩色。多媒体计算机正朝着大存储容量、高运算速度及高品质的视频、音频的方向发展。

MPC 标准是对多媒体计算机的配置规范、必须具备的硬件条件及处理能力的规定和说明。因此，满足或超出 MPC 标准要求的 PC 就可以认为是具有某类特定功能的多媒体个人计算机，而低于 MPC 标准要求的则是非多媒体个人计算机。

MPC 标准的制定对多媒体技术的发展和普及起到了重要的推动作用。随着计算机硬件技术和多媒体技术的飞速发展，MPC 的标准还在不断升级。

1.2.2 多媒体计算机软件系统

多媒体计算机软件系统按功能主要分为系统软件和应用软件，其软件层次结构如图 1.4 所示。

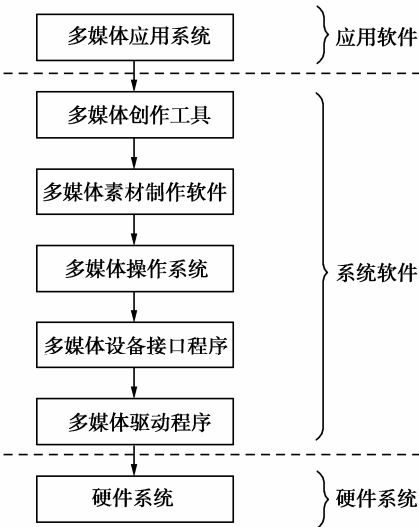


图 1.4 多媒体计算机软件系统结构图

1. 系统软件

系统软件是多媒体系统的核心，主要有以下几种：

(1) 多媒体驱动程序。它是底层硬件的软件支撑环境，直接与计算机硬件相关，完成设备初始化、各种设备操作、设备的打开和关闭、基于硬件的压缩/解压缩、图像快速变换及功能调用等功能。通常，驱动程序有视频子系统、音频子系统及视频/音频信号获取子系统。

(2) 多媒体设备接口程序 (API)。它是高层软件与驱动程序之间的接口软件，为高层软件建立虚拟设备，以便程序员在高层通过接口调用系统功能，并能在应用程序中控制多媒体硬件设备。

(3) 多媒体操作系统。它可实现多媒体环境下的多任务调度，保证音频、视频同步控制及信息处理的实时性，提供多媒体信息的各种基本操作和管理，具有对设备的相对独立性和可操作性。操作系统还具有独立于硬件设备的较强的可扩展性。

多媒体操作系统是多媒体计算机软件系统的核心和基本软件平台，在传统操作系统功能的基础上，增加了处理声音、图像、视频等多媒体功能，并能控制与这些媒体有关的输入、输出设备。目前 PC 上的多媒体操作系统主要有微软公司的 Windows 98, Windows 2000, Windows XP, Windows 2003 等, Apple 公司在 Macintosh 上的 Mac OS 系统, 基于 Linux 的多媒体操作系统, 另外还有早期的 Commodore 公司的 Amiga 操作系统, Philips 和 Sony 公司的 CD-RTOS (CD 实时操作系统) 等。

(4) 多媒体素材制作软件。它是为多媒体应用程序进行数据准备的程序，主要为多媒体数据采集软件，作为开发环境的工具库，供设计者调用。

(5) 多媒体创作工具，又称多媒体著作工具。是指用来集成、处理和统一管理文本、图形、动画、视频图像和声音等多媒体信息的编辑工具。

多媒体应用系统设计不仅要求利用计算机技术将文字、图形、图像、声音、动画及视频等多种媒体有机地融合为图、文、声、形并茂的应用系统，而且要进行精心的创意和精彩的组织，使其变得更加人性化和自然化。多媒体应用系统设计本身很复杂，若用纯编程方法实现，工作量大且难度高，非专业人员莫属，因此多媒体创作工具的研制和推广是十分必要的。

近几年来，各类多媒体创作工具应运而生，如 Macromedia 公司开发的 Director, Authorware; Asymetrix 公司推出的 ToolBook; 由北大方正集团研制的奥斯多媒体创作工具等。多媒体创作工具向管理人员、教育工作者及专业设计者提供了一种途径，即无须成为程序设计专家就可以将不同的媒体融合在一起，它无须掌握深奥的计算机编程技术，面向广大非计算机专业人员，易于学习和使用，可大大缩短开发周期，这种开发方法已被大多数人所接受。

2. 应用软件

应用软件直接面向用户，是为满足用户的各种需求在多媒体创作平台上设计开发的面向应用领域的软件。

1.3 多媒体关键技术

多媒体信息的处理和应用需要一系列相关技术的支持。以下几个方面的关键技术是多媒体研究的热点，也是未来多媒体技术发展的趋势。

1. 大容量数据存储技术

早期的计算机所处理的信息主要是文本文件和数据文件，数据的类型比较单一，数据量也比较有限。随着多媒体技术的应用和普及，各种信息在介质中所占用的空间越来越大，在存储和传输这些信息时需要很大的空间和时间开销，解决这一问题的关键就是数据存储技术。

硬磁盘是计算机重要的存储设备。目前，单个硬盘的容量已达到上百个 GB。磁盘阵列 RAID（Redundant Array of Inexpensive Disk）是由许多台磁盘机或光盘机组成的快速、超大容量外存储系统，如图 1.5（a）所示，其最大集成容量可达上千个 GB 或更多，在一些大型服务器和视频点播系统中广泛采用，是实现高可靠、快响应、大容量存储的必备设备。

光盘的发展速度也很快，如 VCD 采用 MPEG—1 图像压缩技术，已广泛用于电影、卡拉 OK、广告、电子出版物和教育培训等方面，成为市场上最热门的光盘产品之一。DVD（如图 1.5（b）所示）采用 MPEG—2 图像压缩技术，现已推出单面单层、单面双层、双面单层、双面双层 4 种记录密度，其单面单层容量为 4.7GB，而双面双层容量可达 17GB（均指 12cm 盘）。

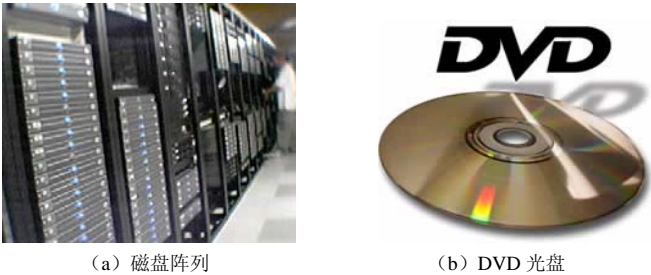


图 1.5 大容量数据存储技术

2. 多媒体数据压缩与编码技术

数字化信息的数据量相当庞大，给存储器的存储容量、通信信道的传输速率（带宽）及计算机的处理速度带来极大的压力。考虑到技术与成本等诸多因素，解决这个问题单纯用增加存储器容量和通信信道的带宽，以及提高计算机的运算速度等办法是不现实的。多媒体数据压缩编码技术（如图 1.6 所示）是解决大数据量存储与传输问题的行之有效的方法。采用先进的压缩编码算法对数字化的视频和音频信息进行压缩，既节省了存储空间，又提高了通信介质的传输效率，同时也使计算机实时处理和播放视频、音频信息成为可能。

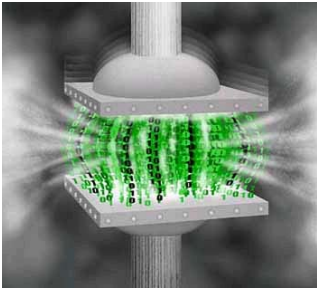


图 1.6 数据压缩示意图

数据压缩技术目前已有以下一些国际标准：

(1) JPEG (Joint Photographic Experts Group) 标准。JPEG 是由国际标准化组织 ISO 等机构联合组成的专家组，专门负责制定静态图像数据压缩的工业标准，其所制定的标准便称为 JPEG 标准。它既可用于灰度图像又可用于彩色图像，由于综合采用多种压缩编码技术，因此经其处理的图像质量高、压缩比大。

(2) MPEG (Moving Picture Experts Group) 标准。MPEG 是负责制定视频和音频编码工业标准的专家组，旨在解决视频图像压缩、音频压缩及多种压缩数据流的复合与同步，它很好地解决了计算机系统对庞大的音像数据的吞吐、传输和存储问题，使影像的质量和音频的效果达到令人满意的程度。

有关音频和图像的压缩算法和标准将在后面的章节里做进一步的介绍。

3. 多媒体通信技术

现代社会人们工作方式的特点具有群体性、交互性。传统的电信业务，如电话、传真等通信方式已不能适应社会的需要，迫切要求通信与多媒体技术相结合，为人们提供更加高效和快捷的沟通途径，如提供多媒体电子邮件、视频会议、远程交互式教学系统、视频点播等新的服务。

多媒体通信是一项综合性技术，涉及多媒体、计算机和通信等领域，它集计算机的交互性、网络的分布性和多媒体信息的多样性于一体，为人类提供了全新的服务。多媒体通信是继电报、电话、传真之后的第 4 代通信手段。

4. 多媒体数据库技术

传统的数据库只能解决数值与字符数据的存储检索。多媒体数据库除要求处理结构化的数据外，还要求处理大量非结构化数据。多媒体数据库需要解决的问题主要有：数据模型、数据压缩/还原、数据库操作、浏览、统计查询及对象的表现。

随着多媒体技术的发展，面向对象技术的成熟和人工智能技术的发展，多媒体数据库、面向对象的数据库及智能化多媒体数据库的发展越来越迅速，它们将进一步发展或取代传统的关系数据库，形成对多媒体数据进行有效管理的新技术。

5. 多媒体信息检索技术

多媒体信息检索是根据用户的要求，对图形、图像、文本、声音、动画等多媒体信息进行检索，得到用户所需的信息。对图像和视频等多媒体信息集来说，目前绝大多数检索系统仍采取文本搜索技术，例如，Google 的图像和视频检索功能仍是基于文本关键词 (Keyword) 的，这些关键词可能来源于图片周围的文本、文件名等，其中少量的也可能来源于人工标注。由于多媒体信息制作者的文化背景不同、专业知识迥异，这些与图片关联的文本信息往往是极不可靠的。对于图像和视频等多媒体信息，一般难以用自然语言进行有效的、精确的描述，无法表达其实质内容和语义关系，所以这种依据文本信息检索图片和视频的解决方案很难满足人们的查询需要，搜索精度很低。

经过 10 多年的努力，许多学者试图实现基于内容的多媒体信息检索技术，以弥补上述多媒体信息检索技术的缺陷。国内外有很多学者在积极研究基于内容的多媒体信息检索技术，其中包括对图像 (如图 1.7 所示)、视频和音频等多媒体信息的内容处理和分析 (Parsing)、

自动标注（Annotation）、构建索引（Indexing）和相似检索（Retrieval）等。

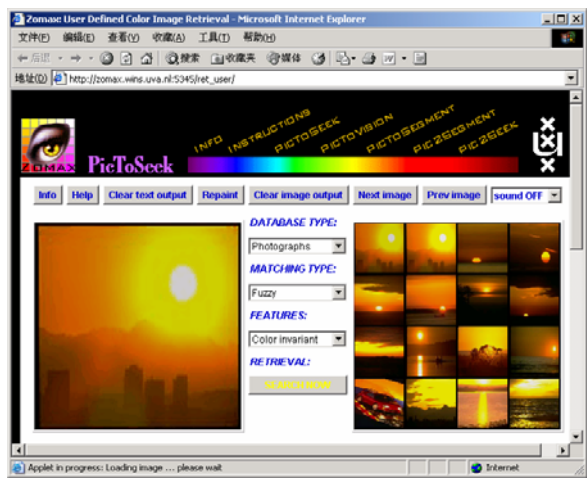


图 1.7 基于内容的图像检索系统——PicToSeek

基于内容的多媒体信息检索技术是一种新兴的信息处理技术，它包含了自然语言处理、图像处理、视频处理、语音识别、数据库与数据挖掘、模式识别、人工智能、机器学习等众多学科，是一个长期的、复杂的、富有挑战性的研究课题。

1.4 多媒体技术的应用及其发展趋势

1.4.1 多媒体技术的应用

以下所述只是其中的几个主要方面。

1. 休闲娱乐

多媒体在娱乐中的应用不仅包括三维游戏（如图 1.8 所示），还包括音乐 CD、VCD、DVD、数字音乐 MIDI 等内容。



图 1.8 三维游戏——极品飞车界面

2. 教育与培训

世界各国的教育学家们正努力研究用先进的多媒体技术改进教学与培训。以多媒体计算

机为核心的现代教育技术使教学手段丰富多彩，使计算机辅助教学（CAI）如虎添翼。

3. 桌面出版（Desktop Publishing）与办公自动化

桌面出版物主要包括印刷品、表格、布告、广告、宣传品、海报、市场图表、蓝图及商品图等。多媒体技术为人类增加了控制信息的能力和充分表达思想的机会，许多应用程序都是为提高工作人员的工作效率而设计的，从而产生了许多新型的办公自动化系统。由于采用了先进的数字影像和多媒体计算机技术，把文件扫描仪、图文传真机、文件资料微缩系统等和通信网络等现代化办公设备综合管理起来，将构成全新的办公自动化系统。

4. 多媒体电子出版物

光盘具有存储容量大、使用和收藏方便、数据不易丢失等特点。它将会在某些领域取代传统的纸质出版物，成为图文并茂的电子出版物。它尤其适合用于各种大容量的出版物，如字典、辞典、百科全书、年鉴、大型画册等。

5. 多媒体通信

在通信工程中的多媒体终端和多媒体通信也是多媒体技术的重要应用领域之一。随着“信息高速公路”的开通，电子邮件已被普遍采用，对人类生活、学习和工作将产生深刻影响的还有信息点播（Information Demand）和计算机协同工作 CSCW（Computer Supported Cooperative Work）系统。

信息点播是指通过桌面多媒体信息系统，人们可以远距离点播所需信息，如视频点播系统（VOD）。

计算机协同工作（CSCW）是指在计算机支持的环境中，一个群体协同工作以完成一项共同的任务，其应用于工业产品的协同设计制造，远程会诊，不同地域位置的同行们进行学术交流，师生间的协同式学习等。

多媒体通信技术不仅改变了信息传递的面貌，带来通信技术的大变革，而且计算机的交互性、通信的分布性和多媒体的多样性的结合，将构成继电报、电话、传真之后的第4代通信手段，向社会提供全新的信息服务。

6. 多媒体作品的创作

多媒体作品包括影片剪接、文本编排、音响、画面等特殊效果的制作等。人们可以通过多媒体创作工具（如 Authorware 等）的帮助增进其作品的品质，MIDI 的数字乐器合成接口可以让设计者利用音乐器材、键盘等合成音响输入，然后进行剪接、编辑，制作出许多特殊效果。还可以用多媒体工具处理视频（如 Premiere 等），制作卡通和动画（如 Flash，3D MAX 等）的特殊效果。

1.4.2 多媒体技术的发展趋势

多媒体技术使计算机具有向人类提供综合声、文、图、像等各种信息服务的能力，从而使计算机进入人类生活的各个领域。分布式多媒体技术又进一步把媒体的多样性与通信的分布性和计算机的交互性相结合，逐渐向人类提供全新的信息服务，使计算机、通信、新闻和娱乐等行业之间的差别正在缩小或消失。多媒体技术正使信息的存储、管理和传输的方式产

生根本性的变化，它影响到相关的每一个行业，同时也产生了一些新的信息行业。其发展趋势有：

(1) 多媒体从单机、单点的研究向分布、协同多媒体网络环境及其设备的研究和网上分布式应用与信息服务的方向发展。

(2) 利用较成熟的图像理解、语音识别、全文检索等技术研究多媒体基于内容的处理，开发能进行基于内容的处理系统（包括编码、创作、表现及应用），是多媒体信息管理发展的重要方向。

(3) 多媒体标准仍是研究的重点。各类标准的研究将有利于产品规范化，应用更方便。因为以多媒体为核心的信息产业突破了单一行业的限制，涉及诸多行业，而多媒体系统的集成特性又对标准化提出很高的要求，所以必须开展标准化研究，它是实现多媒体信息交换和大规模产业化的关键所在。

(4) 多媒体技术将与相邻技术结合以提供完善的人-机交互环境。同时多媒体技术继续向其他领域扩展，使其应用范围进一步扩大。

可以预见，在不久的将来，多媒体将普及到人们工作和生活的方方面面，多媒体技术的发展将是一幅绚丽多彩的画卷。

习 题 1

1. 填空题

- (1) 国际电话电报咨询委员会 CCITT 把媒体分成 5 类：_____、_____、_____、_____、_____。
- (2) 多媒体技术的特征主要有：_____、_____、_____。
- (3) _____是多媒体计算机软件系统的核心和基本软件平台。
- (4) _____是继电报、电话、传真之后的第 4 代通信手段。

2. 简答题

- (1) 什么是媒体？媒体是如何分类的？
- (2) 什么是多媒体技术？它有哪些主要特征？
- (3) 什么是超文本？什么是超媒体？
- (4) 什么是流媒体？
- (5) 简述多媒体计算机系统的组成。
- (6) 简述多媒体所涉及的关键技术。
- (7) 多媒体技术的应用领域主要包括哪些方面？

第2章 数字声音

声音被用来传递消息、思想、情感，是人类最熟悉的传递消息的方式。声音携带的信息量大、精细、准确，是多媒体信息的一个重要组成部分。对声音的合理使用可以使多媒体应用系统变得更加丰富多彩。随着多媒体信息处理技术的发展和计算机数据处理能力的增强，音频处理技术越来越受到重视，并得到了广泛的应用。

2.1 声音信号数字化

2.1.1 声音基础

1. 声音

人们把发出声音的物体称为声源。声源发出声音在空气中引起非常小的压力变化，这种空气的压力变化被耳朵的耳膜所检测，然后产生电信号刺激大脑的听觉神经，从而使人能感觉到声音的存在。所以，自然界中声音是靠空气传播的。各种声音一般都具有周期性强弱变化的特性，因而也使得输出的压力信号周期变化，人们将这种变化用一种图示的方法——正弦波来形象地表示，如图 2.1 所示。

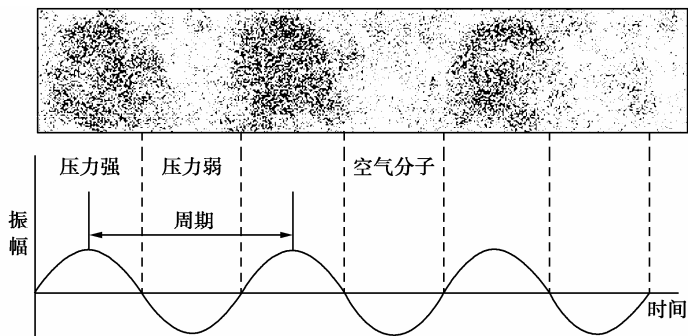


图 2.1 声音的正弦波示意图

人们把曲线上的任一点再次出现（例如，声音由高到低再到高）所需的时间间隔称为周期。而 1 秒钟内声音由高（压力强）到低（压力低）再到高（压力强）这个循环出现的次数称为频率。频率以 Hz（赫兹）为单位，频率越高，声音越高。在实际生活中，声音是混杂的，是由许多不同频率的正弦波合成的复合波。声音的频率范围称为声音的带宽（Bandwidth），它是声音的一项重要参数。多媒体技术处理的声音主要是人耳可听到的 20Hz~20kHz 的音频信号，其中人的说话声音是一种特殊的语音，其频率范围为 300~3400Hz，称为言语（Speech），也称为话音或语音。现实世界中的其他各种声音，如音乐声、风雨声、汽车声等，其带宽要宽得多，可达到 20Hz~20kHz，它们通称为全频带声音。

2. 声音质量分级与带宽

声音的质量与它所占用的频带宽度有关，频带越宽，信号强度的相对变化范围就越大，音响效果也就越好。根据带宽可将声音质量分为如下 4 级：

- (1) 数字激光唱盘质量。通常又称 CD-DA 质量，也就是常说的超高保真，即 Super HiFi (High Fidelity)。
- (2) 调频无线电广播质量，简称 FM (Frequency Modulation) 质量。
- (3) 调幅无线电广播质量，简称 AM (Amplitude Modulation) 质量。
- (4) 电话 (Telephone) 质量。

如图 2.2 所示为这 4 级声音质量的频率范围。在这 4 级质量中，CD-DA 的声音质量等级最高。

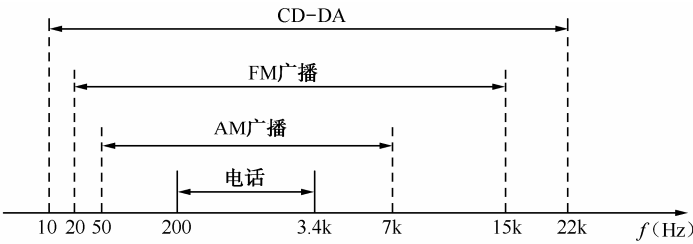


图 2.2 4 级声音质量的频率范围

3. 声音信号数字化

自然界的声音是一种模拟的音频信息，是连续量，而计算机只能处理离散的数字量。如果要用计算机对音频信息进行处理，则首先要将模拟音频信号（如语音、音乐等）转换为数字信号。由于数字化的声音易于用计算机软件处理，所以现在几乎所有的专业化声音录制器、编辑器都是数字的。

2.1.2 声音信号数字化过程

将模拟声音数字化最早采用脉冲编码调制 PCM (Pulse Code Modulation) 技术，它几乎是所有数字音频格式的始祖。1939 年，法国工程师 Alec Reeves 发明了将连续的模拟信号转换成时间和幅度都离散的二进制码代表的脉冲编码调制信号的技术，并申请了专利。PCM 首先开始应用于电话系统，但一直到 1962 年，美国 Bell 实验室才为 AT&T 制成了国际上第一套商用 PCM 电话系统 (T1 系统)，这标志着通信开始进入数字化。以后的计算机发展更促进了通信的数字化，并逐步与通信相结合。由于模拟声音信号非常复杂，PCM 需要通过采样、量化、编码 3 个步骤将连续变化的模拟信号转换为数字信号，其过程如图 2.3 所示。

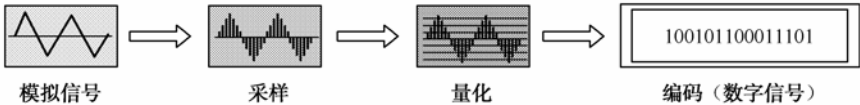


图 2.3 音频信号数字化过程

1. 采样 (Sample)

“采样”就是对模拟信号每隔一个固定的时间间隔取一个样本值。例如，在拍摄电影时，不必要也不可能用胶片把演员的连续动作全部拍摄下来，只是按一定的时间间隔拍摄一幅幅不连续的画面，每幅画面只是取出了演员动作在某一瞬间的“样子”，这就是“采样”。只要采样的时间间隔足够短，如每秒采样不少于 24 幅，那么，在放映时，由于人的眼睛有视觉暂留的特性，在银幕上看到的就不是一幅幅不连续的画面，而是连续的动作了。与此类似，要传输一个声音信号，不必像传送模拟信号那样把信号各个瞬时的声波幅度都连续地传送出去，而只需要每隔一定的时间间隔对信号取一个瞬时值（样本值）传送出去就可以了，如图 2.4 (a) 所示。

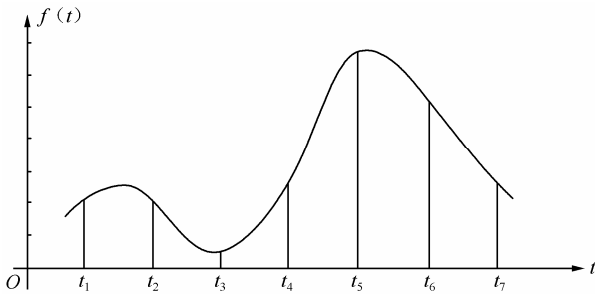


图 2.4 (a) 在离散时间点采样

显然，采样时间间隔越短，所取的一系列值就越能精确地反映原来的模拟信号。否则，采样时间间隔太长，就会使原信号失真。每秒钟采样的次数叫做采样频率 (Sampling Rate)，其单位为 Hz (赫兹)。采样频率越高说明采样时间间隔越短，采样频率越低说明采样时间间隔越长。那么，采样频率要达到多高才能较好地反映出原来的声音呢？

著名的 Nyquist (奈奎斯特) 采样定理认为：采样频率至少应为信号最高频率的两倍，才能把数字信号表示的声音还原为原来的声音。众所周知，人耳的响应频率范围为 20Hz~20kHz，为保证声音不失真，采样频率至少应保证不低于 40kHz。此外，为了避免高于 20kHz 的高频信号干扰采样，在进行采样之前，需要对输入的声音信号进行滤波。考虑到滤波器在 20kHz 处大约有 10% 的衰减，因此再将其提高 10% 成为 22kHz。这个值乘以 2 (双声道) 就得到 44kHz 的采样频率。但是为了能够与电视信号同步，PAL 制式电视的场频为 50Hz (详见第 4 章)，NTSC 制式电视的场频为 60Hz，取 50 和 60 的整数倍，所以 CD 音频通常采用 44.1kHz 的采样频率。与全频带声音相比，人类语音和电话信号的最高频率相对较低，可以选用较小的采样频率。

经过采样后，信号就成为时间上离散的了。

2. 量化 (Quantize)

由于检测器的灵敏度有限和传输中噪声的干扰，没有必要传送一个个样本值的准确大小，只要按照允许的误差将样本大小进行量化分层近似即可。如图 2.4 (b) 所示，将纵坐标上幅度的最大变化范围对应 3 位二进制数码分为 $2^3=8$ 个层级，即 0, 1, 2, 3, 4, 5, 6, 7。然后用四舍五入的方法将样本大小用其接近的分层级来代替，这个过程叫做“量化”。图中对应于时刻 $t_1, t_2, t_3, t_4, t_5, t_6, t_7$ 的样本值分别代替为 2, 2, 0, 3, 7, 6, 3。模拟信号经过

采样、量化，在时间、幅度上都离散化、不连续了。

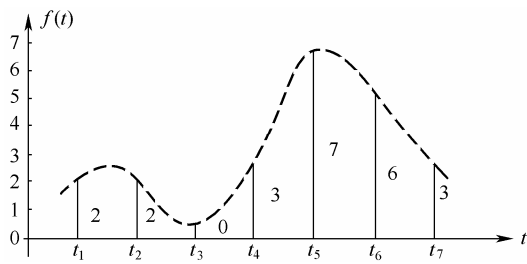


图 2.4 (b) 对样本值大小进行量化

显然，所分的层数越多，即二进制码的位数越多，则对原模拟信号反映的精度就越高，对于声音信号来说，记录的声音质量就越好。采用 3 位二进制编码，只分为 2^3 个层级，其精度只有 $1/2^3$ ，即 $1/8$ ；如用 8 位二进制编码，则可分 2^8 个层级，精度可达 $1/2^8$ ，即 $1/256$ 。而用 16 位记录的声音质量又比 8 位记录的声音质量好得多。当然，二进制码位数的多少受到技术上的限制，常用 8 位、12 位、16 位、24 位、32 位等。同理，在相同量化位数的情况下，采样频率越高，声音效果也越好。

3. 编码 (Code)

采样、量化后的信号还不是数字信号，需要把它转换成数字编码脉冲，这一过程称为编码。最简单的编码方式是二进制编码，如图 2.4 (c) 所示的样本值 2, 2, 0, 3, 7, 6, 3 用二进制数码表示就是 010, 010, 000, 011, 111, 110, 011，并按一定规则顺序排列形成脉冲序列，这就是“编码”。这样就生成了以数字信号表示的信息。在实际过程中量化和编码是同时进行的。

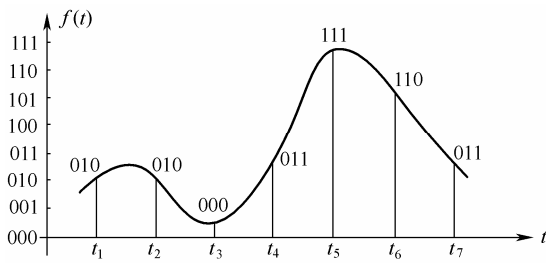


图 2.4 (c) 对样本进行编码

2.1.3 声音数字化的主要参数

音频数字化技术是整个数字音频领域中最基本、最主要的技术。在微机中，这一过程是由声音处理卡 (Sound Card) 及相关软件完成的。声音处理卡上的 A/D 转换电路对输入的音频模拟信号以固定的时间间隔进行采样，然后将采样信号送给量化编码器，变成数码，并将所得的数码保存下来。每一个采样值对应于音频信号某一点的瞬时值。

音频数字化有 3 个重要参数：采样频率 (Sample Rate)、量化精度 (Resolution)、声道数 (Channels)。下面分别介绍 3 个参数对音频数字化的影响。

1. 采样频率

采样频率是指在 1 秒钟时间内采样的次数。采样频率越高，声音保真度就越高，产生数据量也越多，所需的存储空间也越大。音频采样频率的常见标准有 11.025kHz，22.05kHz，44.1kHz，48kHz 等。

2. 量化精度

量化精度是指每个采样点所表示的数据位数。不同的位数（精度）决定了不同的音质，位数越多，精度越高，对原始波形的模拟就越细腻，失真度也就越小。同时，需存储数据量也越大。量化精度通常有 8 位、16 位、24 位。

3. 声道数

古人说“听音辨位”，人可以感觉到声音的来源方位，还可以感觉声音的强、弱、远、近，可是当声音可以通过电子设备传递后，声音和距离已经没有什么绝对关系了，因为讲话的人可能远在千里之外。随着科学技术的不断进步，人们要求声音要有现场感，希望感受到的声音就好像置身于现场听到的一样，于是声道技术开始发展。简单地说，声道（Channel）是指同一时间有几个喇叭发出声音。在这里声道数指同一时间中出现（产生）的音频通道数。常见的有单声道和双声道（立体声），立体声波形数字化处理后的数据存储量是单声道的 2 倍。单声道缺乏对声音的位置定位，而立体声在录制过程中被分配到两个独立的声道，从而能达到很好的声音定位效果，现场真实感强。要让声音得到真实的还原则应尽量采用立体声录制和重放。现在的 FM 广播电台基本上都是立体声广播。



图 2.5 5.1 声道的音箱系统

立体声虽然满足了人们对左、右声道位置感体验的要求，但要达到好的效果，仅仅依靠两个音箱是远远不够的。随着声音合成技术的发展，双声道立体声逐步向多声道环绕声发展。早期运用了 3 个喇叭（两前一后）来达成，通常是配合电影的画面效果来模拟的，例如，画面中一架飞机从远处向你飞来，引擎呼啸声由远而近，让你感觉就好像身处飞机场一样。往后便发展到 4 声道（两前两后）、DVD 的 6 声道（5.1）、7 声道（6.1）、8 声道（7.1），以及电影院的 10 声道。5.1 声道的音箱系统如图 2.5 所示。

在一向讲究软硬兼施的 PC 界，多声道音频的实现自然少不了软件算法的控制过程。因此，3D 音频 API 就扮演了重要的接口角色。当然，有些 API 还包含了具体算法。这些 API 与 3D 图形程序接口，统称为 3D API（Application Program Interface），即 3D 应用程序接口。对于支持 3D 定位技术的新一代声卡而言，算法往往决定了其定位及其他效果的优劣，因此，我们有必要对它们的编程接口有一些简单的了解。

音频 API 种类繁多，目前各种游戏可以使用的 API 和 3D 技术大体上有 DirectSound 3D（DS3D），A3D，EAX，Sensaura 3D，Q3D，IAS 等。不同的声卡硬件和不同的游戏往往支持多种不同的 API 和 3D 技术，这主要取决于声卡所采用的音效芯片的类型。

（1）DirectSound 3D

DirectSound 3D（即 DS3D）是 DirectX 中的一个组件，是 Microsoft 公司专为游戏开发的 API，得益于 DirectX 的不断发展和完善，DS3D 得到了众多声卡厂商的支持。DS3D 的作用

在于帮助开发者定义声音在 3D 空间中的定位和声响，然后把它交给与 DS3D 兼容的声卡，让它们用各种算法去实现。定位声音的效果好坏实际上取决于声卡所采用的算法。

（2）Aureal 3D

Aureal 3D 简称 A3D，是由著名的 Aureal Semiconductor 公司开发的新型 3D 音效定位技术。使用这一技术的应用程序（通常是游戏）可以根据用户的选择来决定音效的变化，而且可以只通过一对普通的音箱或耳机来实现，产生围绕听者的 3D 精确定位音效。

Aureal 推出的 3D 定位音效标准 A3D 3.0 强化了所有 A3D 2.0 的功能，如音源定位、声波轨迹（WaveTracing）、回声处理、封闭空间音效等。此外它还加入了几项新的功能，如支持杜比数码音频（Dolby Digital，即 AC3）解码、回声几何运算功能（Geometry Engine）、大型音源效果（Volumetric Sources）处理等。

但很可惜，A3D 技术最终未能逃过被终结的命运。在 2000 年 8 月，Aureal 被创新科技公司吞并，A3D 技术从此退出音频舞台。A3D 技术无论是在当时，还是在今天都是一种技术非常优秀的 3D 音频 API。

（3）EAX

创新公司推出了自己的首个三维音效技术——EAX（Environmental Audio Extensions，环境音效扩展）。EAX 环境音效技术是当时唯一一个可与 A3D 相抗衡的三维音频技术。就技术而言，EAX 并不是一套独立的 API，而是微软 DirectSound 3D 的加强功能。EAX 环境音效技术不但能把声音变换成各种各样的特效声，而且还能把来自 CD，VCD，MIDI，WAV 或 MP3 等立体声音源虚拟为多声道环绕音效。另外它的全方位音频定位技术也是一绝。现在，EAX 技术已经占据了绝大部分游戏的三维音频系统。

（4）Sensaura

来自英国的 Sensaura 创造研发出多种 3D 音效定位的演算法，并且授权给音频芯片开发与音响硬件公司。使用 Sensaura 技术的产品一般都能兼容于 A3D 与 EAX，而且一般都包含了数项 DirectSound 3D 的加强功能。当然 Sensaura 拥有自己特有的技术，如 MacroFX，它是一项可以模拟非常贴近使用者头部的真实音源技术；ZoomFX，允许设计师创造定位出大型的发声来源；EnvironmentFX，效果类似而且共通于 EAX 的效果，它能够处理不同环境下的 3D 音场，如水面底下的环境、大型的房间、金属的墙壁等。Sensaura 后来被创新收购后，并未像 A3D 一样从此消失，其技术仍然得到广泛的应用，现在它主要被用于电视游戏机及一些专业应用软件和硬件设备上。

（5）Qsound

与 Sensaura 一样，Qsound 的 Q3D 也是相当古老的 3D 音效 API，自从 1991 年进入市场以来，经过了几次大的变革，已成为一种十分成熟的 3D 音效。不过 Qsound 的三维音频技术在 PC 平台上并没有取得成功。幸好 Qsound 的 3D 音频技术不仅用在娱乐声卡领域，所以它并没有像 A3D、Sensaura 那样在 PC 平台上经营失败后被收购。其实，它在游戏机方面的应用比 PC 更为广泛，在许多大型游戏机中都可以看到 Qsound 的身影。

（6）IAS

上面这么多的 API 和技术，它们各有特点，这样就必须针对不同的系统和 API 编写多套代码，IAS（Interactive Around-Sound）就是针对这个特点而形成的。

IAS 是 Extreme Audio Reality Inc 公司开发的专利音频技术，这个技术能测试系统硬件，管理所有的音效平台，因而开发者可以只写一套音效代码，所有基于 Windows 的音频硬件

将通过同样的编程界面来获得支持。IAS 提供了 DS3D 支持和其他环绕声的执行程序。

2.1.4 数字化声音的压缩

1. 声音的码率和声音文件大小

声音每秒的数据量称为声音的码率。码率也称比特率，声音的码率就是每秒记录音频数据所需要的比特值，通常以 kbps（千比特/秒）为单位。CD 中的数字音乐码率为 1411.2kbps，而 MP3 音频的码率在 112~128kbps 时即可实现接近于 CD 的音质。声音未经压缩时的码率可由下式算出：

$$\text{声音的码率} = \text{采样频率} \times \text{量化精度} \times \text{声道数}$$

表 2.1 是电话中继线和长途线路上传输的数字语音和 CD 唱片上高保真、全频带立体声数字声音的主要参数。

表 2.1 两种常用数字声音的主要参数

声音类型	声音带宽 (Hz)	采样频率 (kHz)	量化位数 (bits)	声道数	未压缩时的码率 (kbps)
数字语音	300~3400	8	8	1	64
CD 唱片	20~20 000	44.1	16	2	1411.2

声音数据量以 Byte 为单位，可由下式算出：

$$\begin{aligned} \text{声音数据量} &= \text{采样频率} \times (\text{量化精度} \div 8) \times \text{声道数} \times \text{时间} \\ &= (\text{声音的码率} \div 8) \times \text{时间} \end{aligned}$$

例如，CD 唱片的采样频率是 44.1kHz，量化精度为 16 位，声道数为 2（立体声），那么，根据上式计算出每分钟声音的数据量为 $44\,100 \times (16/8) \times 60 \times 2 \approx 10.09\text{MB}$ 。

由此可看出，取样频率和量化精度越高，声道数越多，所需的存储空间也就越大。也就是说，声音的码率越大，所需的存储空间也就越大。当然数据量大小还跟保存声音的时间有关。

2. 数字化声音的压缩

存储 1 分钟采样率为 44.1kHz、量化级为 16 位、双声道的 PCM 编码的音频信号，需要 10.09MB 的空间，这显然太庞大了。要减少磁盘占用有两种方法，一种方法就是降低采样率，但降低采样率会严重影响音质，因而是不可取的；另一种方法是对声音数据进行压缩。

为了对声音数据进行压缩，出现了各种压缩算法。它们可以分为有损压缩与无损压缩，实际上有损压缩和无损压缩也只是相对的。音频编码最多只能做到无限接近于自然界的信号，实际上，任何数字音频编码都无法做到完全还原自然的的声音信号。在所有的数字音频编码中，PCM 编码代表了最高的保真水平，因此，它被约定成为无损编码。而通常上讲的有损音频编码，只是相对于 PCM 编码而言的。

压缩率通常指音乐文件压缩前和压缩后大小的比值，用来简单描述数字声音的压缩效率。

2.1.5 声音文件和格式

到目前为止，出现过的数字音频格式已经数不胜数了，有些还在广泛的使用，有些已经没有生命力了。这里我们介绍几种常见的音频格式，并根据它们出现的时间及存在的方式进

行了简单的分类。

1. 早期数字音频格式

(1) PCM

PCM 即 Pulse Code Modulation (脉冲编码调制) 的缩写, 它可以算做数字音频编码的鼻祖。它的最大优点就是保真度高, 缺点就是体积庞大。常见的 Audio CD 就采用了 PCM 编码, 一张光盘的容量只能容纳 72 分钟的音乐信息。PCM 一般作为 WAV 文件的编码而存在, 也有以 PCM 为扩展名的音频文件, 不过比较少见。而最新的 DVD-Audio 也使用 PCM 编码, 其指标相当高, 采样率最高可达 192kHz, 量化级为 24 位。

(2) WAV

WAV 是由微软开发的一种古老的音频文件格式, WAV 格式是非常灵活的, 它对音频流的编码没有硬性规定, 除了 PCM 之外, 几乎所有支持 ACM (Audio Compression Manager) 规范的编码都可以为 WAV 的音频流进行编码, 如 ADPCM, GSM, CCITT, G.723.1 等, 甚至也包括 MP3 格式。虽然 WAV 文件可以存放压缩音频甚至 MP3, 但由于它本身的结构注定了它的用途是存放音频数据并用于进一步的处理, 而不是像 MP3 那样用于聆听。目前所有的音频播放软件和编辑软件都支持这一格式, 并将该格式作为默认文件保存格式之一。这些软件包括 Sound Forge, Cool Edit Pro, WaveLab 等。

(3) AIFF/AU

AIFF 是苹果电脑中的标准音频格式, 属于 QuickTime 技术的一部分。它和 WAV 非常类似, 实际上 WAV 格式就是由它克隆而来的。不过 AIFF 远不如 WAV 盛行, 但由于苹果电脑在多媒体领域里的领导地位, 所以, 大部分的音频编辑软件和播放软件都对它提供了支持。AU 则是 UNIX 平台下一种常用的音频格式, 起源于 Sun 公司的 Solaris 系统。AU 格式本身也支持多种压缩方式, 但其文件结构的灵活性就比不上 AIFF 和 WAV 了。由于 UNIX 平台比 Mac 平台更为少见, 了解它的人并不多, 因而, 它得到的支持也远不如 AIFF 和 WAV。

(4) MIDI

MIDI 文件本身只是一堆数字信号而已, 不包含任何声音信息, 它记录的是在音乐的什么时间用什么音色发多长的音等。正因为这样, 通常的 MIDI 文件个头都非常小。由于 MIDI 记录的并不是真正的声音, 不同的声卡、软波表、硬件音源的音色是不相同的, 所以相同的 MIDI 文件在不同的设备上播放会有不同的效果。普通的 MIDI 文件许多播放器都支持, 但要达到好的效果就必须安装软波表。目前 Microsoft 已经在 DirectX 里面增加了软波表, 并使之成为 Windows 系统的标准配置。MIDI 文件的编辑软件主要有 Cake Walk Pro、Sonar 及国产的 TT 作曲家, MIDI 在音乐创作中有着非常广泛的应用。

2. 新型数字音频格式

(1) MP3

MP3 的名声之响绝对超过其他任何一种音频格式, 因为它的应用范围早已经不仅限于计算机了。MP3 是 MPEG—1 Audio Layer 3 的简称, 属于 MPEG—1 的派生音频方案。MP3 采用的是一种有损压缩方式, 虽然并不是第一个进行有损压缩的音频编码, 但它却是第一个实用的有损音频压缩方案。在 MP3 之前, 一般的音频编码即使以有损方式进行压缩最多达到 4:1 的压缩比, 而 MP3 可以实现 12:1 的压缩比。MP3 使用的是知觉音频编码技术, 它利用了

人耳的特性，削减音频中人耳听不到的成分，同时尽可能地维持原来的声音质量，这是它实现高压缩比，保持较高音质的一个关键。事实上，MP3 在高码率时其音质与 WMA 这类音频格式相比，依然有着较明显的优势。

（2）Real Audio（RA）

Real Audio 属于 Real Media 的音频部分，它是流媒体行业中的霸主 RealNetworks 的产品，它采用流式传输方式，可以在非常低的带宽下（低达 28.8kbps）提供足够好的音质让用户能在线聆听。不过，由于 Real Media 是从极差的网络环境下发展过来的，所以 Real Audio 的音质并不怎么样，在高码率时它要差于 MP3。尽管后来 RealNetworks 通过与 SONY 公司合作，利用 SONY 的 ATRAC 技术实现了高比特率的高保真压缩，但这些已经无法改变它在用户心目中音质差的印象。这也是为什么许多音乐网站能够提供免费的 RA 音乐下载的一个重要原因，音质较差的音乐不用太多担心版权问题。而且 RA 音乐的主要用途是在线聆听，它不适用于编辑，编辑处理软件也非常少。

（3）Windows Media Audio（WMA）

为了挑战 RealNetworks 在流媒体领域的霸主地位，微软推出了 Windows Media 与其相抗衡，Windows Media Audio 即 Windows Media 的音频部分。最初的 Windows Media Audio 并不比 Real Audio 强多少，但随着 Windows Media 9 技术的推出，Windows Media Audio 已经可以令人刮目相看了。微软声称，在只有 64kbps 的码率情况下，WMA 可以达到接近 CD 的音质，Windows Media Audio 的压缩技术中还拥有可变码率（VBR）、无损压缩技术，并支持多声道编码。WMA 还加入了 DRM（Digital Right Management，数字版权管理）技术，可防止复制以保护版权。

3. 常见的视频伴音格式

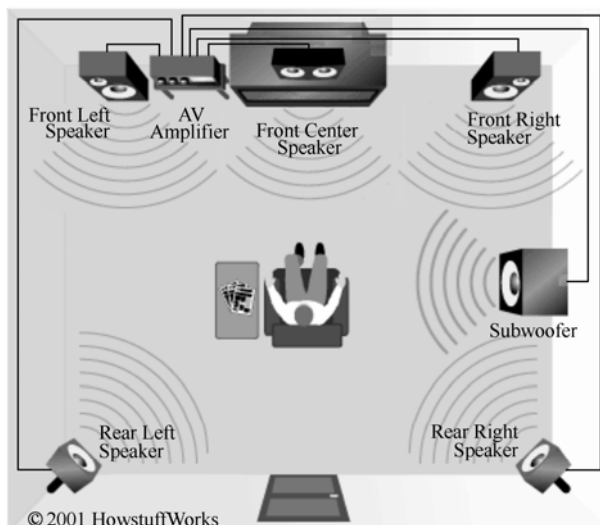
这一部分介绍的音频格式主要是作为电影伴音而存在的，它们很少单独存在，但我们和它们接触的机会并不会比以上两大类少。

（1）Dolby Digital AC3

大部分人是通过 DVD 才了解到 AC3 的，实际上它最初出现是为了在电影院中应用的。它是由 Dolby 公司在 1992 年提出的，最初被称为“杜比 AC-3（Dolby Surround Audio Coding-3）”，以后又改为“杜比数码环绕声”（Dolby Surround Digital），我们通常称它为 Dolby Digital AC3。

实际上 AC3 有多个标准，从单声道、双声道、四声道、五声道一直到我们所熟知的 5.1 声道（Dolby Digital 5.1），它的压缩率最大约为 10：1。Dolby Digital 5.1 提供的环绕声系统由 5 个全频域声道和 1 个超低音声道组成，被称为 5.1 声道，如图 2.6 所示。

5 个声道包括左前（L）、中央（C）、右前（R）、左后（LS）、右后（RS）声道。低音声道主要提供一些额外的低音信息，使一些场景，如爆炸、撞击等声音效果更好。由于每个声道都是用独立麦克风来录制的，所以 AC3 拥有完整的定位能力。AC3 数据的码率，两声道是 192kbps，5.1 声道的码率是 384~448kbps，最高为 640kbps。AC3 早期广泛的一个用途就是作为 DVD 的伴音，带 AC3 音频的 DVD 影片占据了市场上的绝大部分。5.1 声道的 AC3 音频在 PC 上完美回放需要拥有六声道的声卡/音箱，也可通过 S/PDIF（Sony/Philips Digital Interface）接口输出到家庭影院系统中播放，否则，它和双声道音频没有太大的区别。



Dolby Digital 5.1 Installation
2x front stereo speakers 2x rear stereo speakers
1x center speakers 1x subwoofer

图 2.6 Dolby Digital AC3 5.1 声道系统

(2) DTS

DTS 即 Digital Theater Systems（数字影院系统）的简称，它采用了与 AC3 不同的算法，但可以兼容为 AC3 所搭建的音响系统。从音质上看，DTS 要优于 AC3，DTS 在 DVD 上拥有 1536kbps 的数据流量，比 AC3 常用的 384~448kbps 足足多了 3 倍多，这使得 DTS 在声音的细节表现、空间感方面更加优秀，不过目前它还不如 AC3 流行。

2.2 声卡

声卡（也称为声效卡）在多媒体计算机中，是不可缺少的重要部件。它直接决定了多媒体个人计算机 MPC 对声音数据的处理能力与质量。现在的声卡已不仅作为发声之用，还具备声音的采集、编辑、语音识别、网络电话等功能。

2.2.1 声卡介绍

一块声卡（如图 2.7 所示）有哪些作用呢？平时我们除了播放 MP3/CD 以外，还有什么地方会用到声卡呢？其实声卡不仅是一块能够发声的卡，还是一块功能众多的多功能卡。

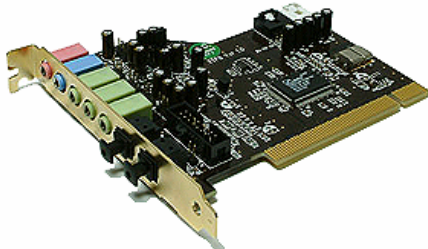


图 2.7 声卡

1. 声卡的功能

(1) 播放数字音乐

这是声卡最基本的功能，这得益于数字音乐的存储方式的改进。从原始的 WAV 到流行的 MP3，再到新兴的 WMA 等音频格式，使得数字音乐被广大用户接受。最初的声卡仅能播放一点简单的提示音，到后来的 8 位声卡、16 位声卡，而目前高档的声卡更是将声卡支持的最大采样速率推向 192kHz，量化精度也高达 24 位。因此我们能够在计算机上欣赏 MP3/CD 甚至 DVD Audio。

(2) 录音

这也是声卡最基本的功能之一，将外部的声音模拟信号，通过声卡录入计算机，实现从模拟信号到数字信号的转换，并以数字化声音文件的形式进行保存。录制的声源可以是麦克风、线路或 CD 输入。目前高档声卡支持 96kHz、24 位的录音，实际录音效果也达到了很高的水准，这样为组建家庭录音室的硬件环境提供了更好的保障。

(3) 实时的效果器

当电脑游戏越来越 3D 化的时候，用户不但要求画面够真实，也要求声音也能够尽量模拟真实环境。有厂家提出了 3D 音效的方案，其中最著名的是 Aureal 的 A3D、Microsoft DirectSound 和 Creative EAX。这些方案的提出，要求声卡有非常强大的运算能力以模拟真实的音效。这些功能的实现和声卡处理信号的能力有关，不但在游戏中需要这些音效，在音乐欣赏的时候也同样需要。著名的 Winamp 播放器就有很多 DirectSound 的实时效果器，在听音乐的时候，我们可以加入回音、混响的效果，这些效果的实现也需要声卡加速来支持。当然，这些效果的实现不一定需要声卡，也可以通过软件模拟，但硬件加速不会牺牲性能。

(4) 音频解码

在数字音频世界中，存在各种各样的编码方案，大部分都可以由软件解决。当 DVD-ROM 开始普及时，人们发现 DVD 对系统的要求比 VCD 要高得多，尤其是音频部分。为了更好地做到身临其境，DVD 影片大部分都采用了 AC3/DTS 方式编码其音频部分，算法远比以前大部分音频编码要复杂，软件也可以解码，但许多人更信赖硬件解码。Audigy2 ZS 声卡允许用户在播放普通的 DTS 音效的 DVD 时，将 5.1 矩阵解码成 6.1，这些都是声卡在起作用。另外，例如，CS4630 系列的声卡，大部分都支持硬件 MP3 加速，其实这就是声卡在参与解码的结果。

(5) 音乐合成

播放 MIDI 的效果的好坏取决于声卡的合成能力的好坏。声卡波表合成能力的好坏直接影响到声卡播放 MIDI 的性能。波表的英文名称为“WAVE TABLE”，从字面翻译就是“波形表格”的意思。其实它是将各种真实乐器所能发出的所有声音（包括各个音域、声调）录制下来，存储为一个波表文件。播放时，根据 MIDI 文件记录的乐曲信息向波表发出指令，从“表格”中逐一找出对应的声音信息，经过合成、加工后回放出来，这个能力的好坏是声卡播放 MIDI 的重要性能指标。

2. 声卡的分类

声卡的分类可以采用多种方式划分。如依据数据采样位数确定，可以划分为 8 位声卡、16 位声卡、24 位声卡，位数越多，音质越好。目前 16/24 位卡占主流，8 位卡被淘汰。如采

用总线方式来划分,可分为 ISA 声卡、PCI 声卡、USB 声卡。由于 PCI 总线的优越性,PCI 声卡有许多 ISA 声卡无法拥有的特性,但这并不是说 PCI 声卡的音质一定比 ISA 好,决定音质的好坏主要靠声音处理芯片、MIDI 的合成方式和制造工艺等,并不仅仅由总线的不同来决定。随着 USB 的全面普及,USB 声卡在市场上也随着这股势头如雨后春笋般地冒出来。这种声卡类型首先在移动性上更加便携,还由于 USB2.0 的普及使得播放多音频流的性能也不会出现采用 USB1.1 接口时的传输瓶颈。另外 USB 声卡由于整个设备是外置的,所以可以完全避免计算机机箱内部杂乱的电磁干扰,可以达到较高的信噪比,因此不同用户可根据自己的环境、实际情况选用 PCI 还是 USB 的独立声卡产品。如按照与计算机的连接方式来划分,可分为独立声卡和集成声卡,后者是集成在计算机主板上的。通常在抗干扰能力、声音处理效果和功能种类等方面,集成声卡都不如独立声卡。如按照用途来划分,又可分为普通声卡和专业声卡。专业声卡,指的是专为录音与电脑音乐制作而设计的声卡产品,通常业内人士也将其称为“音频卡”或“音频工作站”,是相对于那些以电脑游戏和家庭娱乐为主要用途的普通多媒体声卡来说的。

3. 声卡的性能指标

声卡的性能指标决定了声卡声音采集、合成与播放的质量,主要取决于以下几个方面。

(1) 采样和量化能力

主流声卡分为 11.025kHz, 22.05kHz, 44.1kHz, 48kHz 等几个采样等级。量化精度,常见的有 8 位、16 位、24 位。采样频率、量化位数越高,声卡产生的声音就越细腻。

(2) 芯片类型

采用什么样的核心器件是决定声卡性能高低的主要因素。有些声卡采用的芯片性能比较差,有些控制任务要由 CPU 完成,这样的声卡价格较为便宜。声卡专用的数字信号处理器(DSP)集成度很高,不但具有数字信号处理能力,而且还集成了 A/D 转换器,甚至集成了音乐合成器。这样的芯片处理能力很强,对 CPU 的依赖性小,因此采用这种芯片的声卡性能好,但是价格较高。

(3) 输出声道数

声卡所支持的声道数的增加也是声卡技术发展的重要标志之一,它决定了声卡的基本功能。通常有 2 声道(即立体声)、2.1 声道、5.1 声道等。多通道声卡是营造逼真音效环境的先决条件。

(4) 合成技术

现在的声卡都支持 MIDI (Musical Instrument Digital Interface) 标准。MIDI 是电子乐器的统一标准。声卡中一般采用两种不同的方法来还原 MIDI 声音。FM 是一种用计算机合成音调模拟乐器曲调的技术,这种技术已经基本过时了。波表技术要比 FM 合成出色,因为声卡不是用计算机的声调去合成音乐,而是在一个波表(一种内部固有的实际录音选择表)中找到它需要的乐器声音。波表技术能比 FM 合成创作出更好、更自然的声音。

(5) 兼容性

开发、生产声卡的公司很多,其中最有影响的公司是新加坡创新科技有限公司(Creative Labs. Inc.),它开发了 Sound Blaster 系列声卡,它是集语音与音乐于一体的多媒体音频卡,它不但具有优良稳定的硬件特性,而且还有丰富的软件。尽管目前世界各国开发了很多品牌的声卡,但大多都声明与 Sound Blaster 兼容,因此它已成为多媒体计算机公认的音频接口标

准。当然，还有其他一些声卡标准，如 AdLib、Disney Sound Source（迪斯尼声源）等，在选购时，应确保应用的软件支持所选购的声卡，否则可能出现不兼容。例如，如果要使用 Malt Disney 的游戏节目和教育软件，那么必须使用那种完全与众不同的声卡：迪斯尼声源。

【例 2.1】 创新 Sound Blaster Audigy LS 声卡性能参数。

创新 Sound Blaster Audigy LS 声卡性能参数如图 2.8 所示。



图 2.8 创新 Sound Blaster Audigy LS 声卡性能参数

2.2.2 声卡的工作原理

声卡采用大规模集成电路设计，即将音频处理技术以专用芯片形式集成在声卡上，并可直接插入计算机的扩展槽中使用。虽然声卡的品牌与型号各异，功能也不尽相同，但基本包含以下各功能及部件，如图 2.9 所示。

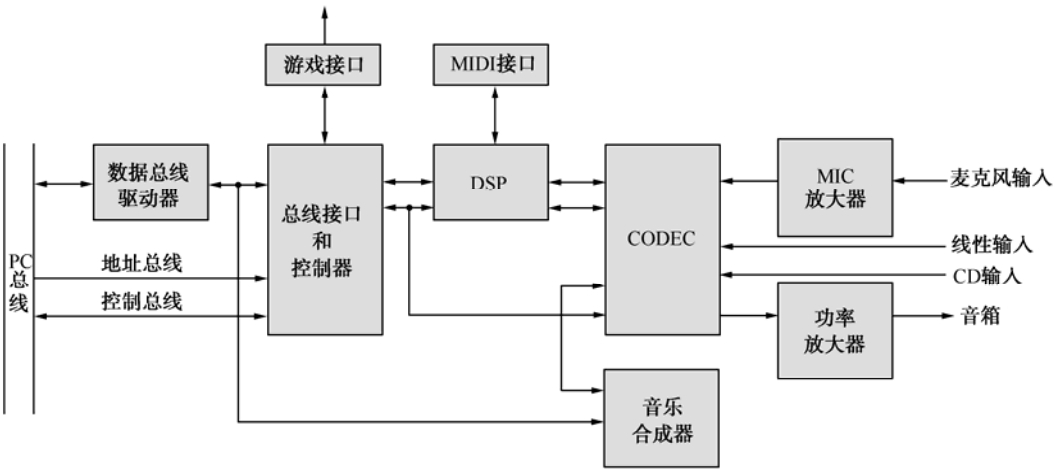


图 2.9 声卡工作原理框图

1. 主芯片

声卡的主芯片承担着三维音效处理的运算和加速、MP3 解码、MIDI 的合成及硬件等级的声音处理（如硬件 EQ 均衡器、各类混响处理）等重要工作。而对声音信号的采样与编码则被分离出来，这个任务交给一个被称为 CODEC 的芯片来完成。

声卡的主芯片普遍都是一块具有强大运算能力的 DSP（Digital Signal Processor，数字信号处理器），是声卡上的一个大块头的芯片。DSP 不是很规范的称呼，正规的场合一般称为 Audio Accelerator，即音频加速器。音频加速器有强大的运算能力，专门为音频处理服务。并不是所有的声卡都有 DSP 芯片，随着技术的发展，DSP 功能完全可以让 CPU 代劳。和 CPU 一样，不同音频加速器有着处理能力上的区别，可以用 MIPS（Million Instructions Per Second）表示运算能力的大小，MIPS 表示一秒执行一百万条指令的能力。我们常见的 Creative Sound Blaster Live 系列采用的音频加速器为 Emu10K1（如图 2.10 所示），拥有 1000MIPS 的运算能力。正因为有了这些音频加速器，我们才能以不牺牲性能来欣赏逼真的游戏音效。



图 2.10 Creative Emu10K1 芯片

2. CODEC芯片

CODEC 是由两个英文字的一部分所组成的，它是 Coder 与 Decoder 组合而成的缩写字，直译为编码器与解码器，完整的称呼是“多媒体数字信号编解码器”，一般我们把它简称为“混音芯片”。说到 CODEC 芯片就必须提到 Intel 公司制定的 AC'97 音频规范，在这份文件中建议为了提高声音信号转换过程中的信噪比，减少电磁干扰，应该把数/模转换（D/A）和模/数转换（A/D）部分从主芯片中脱离出来，采用一个独立的处理单元来进行声音采样和编码，CODEC 也就应运而生了。它一般是一块 48pin 或者 64pin 的小芯片，相对主芯片来说它并不太起眼。CODEC 芯片如图 2.11 所示。

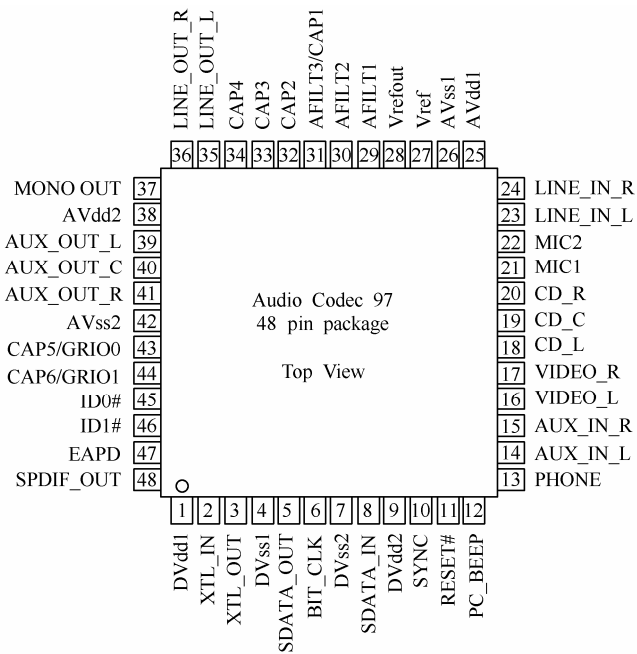


图 2.11 CODEC 芯片

当然 CODEC 技术只是 Intel 的建议，而非强制的标准，例如采用 C-Media 公司 CMI8738 系列芯片的声卡就没有独立的 CODEC 芯片，原因在于 CMI8738 及前几代产品一直都采用单

芯片解决方案，D/A 和 A/D 内置在主芯片内。这样做的好处是简化了声卡的 PCB 设计，降低了总体成本，但不利于提高声卡的信噪比。CODEC 技术成熟以后，板载 AC'97 软声卡也就诞生了，在主板上集成一块混音芯片，而将除了信号采样编码之外的各种声音处理都交由 CPU 来完成运算，以牺牲系统资源和很多附带功能，来换取性价比。

3. I/O控制芯片

I/O 控制芯片，顾名思义就是负责和协调声卡输入/输出控制的。大多数声卡将 DSP 和 I/O 控制芯片集成在一起。音频加速器集成的 I/O 控制器显得过于简陋，无法胜任某些较为专业的场合，例如，需要用到 ASIO，GSIF 的软件，以及多路的 I/O 控制，还有要达到更高指标的输入/输出要求。单独的 I/O 控制芯片就为此而诞生，由于有着更为强大丰富的 I/O 控制能力，因此可以为改善音质做出巨大贡献。

4. 音乐合成器

标准多媒体 PC 可以通过声卡的内部合成器或主机 MIDI 端口的外部合成器播放 MIDI 文件。MIDI 合成器有两种：频率调制（FM）合成器、波表（Wavetable）合成器。波表合成器是将每一种乐器对应一种或几种波形，在合成音乐时，以查表的方式来获取乐器的波形，从而产生效果逼真的合成音乐输出。

5. I/O设备接口

声卡外部输入/输出口均为 3.5mm 规格插口，比较常见的包括如下几种插口。

- ① 麦克风接口（MIC IN）：连接麦克风，实现声音输入、外部录音功能。
- ② 线性输入口（LINE IN）：连接各种音频设备的模拟输出，实现相关设备的音源输入。
- ③ 音频输出口（LINE OUT）：连接多媒体有源音箱，实现声音输出。
- ④ 扬声器输出（SPK OUT）：通过声卡功放输出的放大信号，用于连接无源音箱。
- ⑤ MIDI 设备接口/游戏手柄接口（MIDI/Joystick）：连接 MIDI 音源、电子琴或者游戏控制设备。

某 16 位声卡的接口与设备连接如图 2.12 所示。

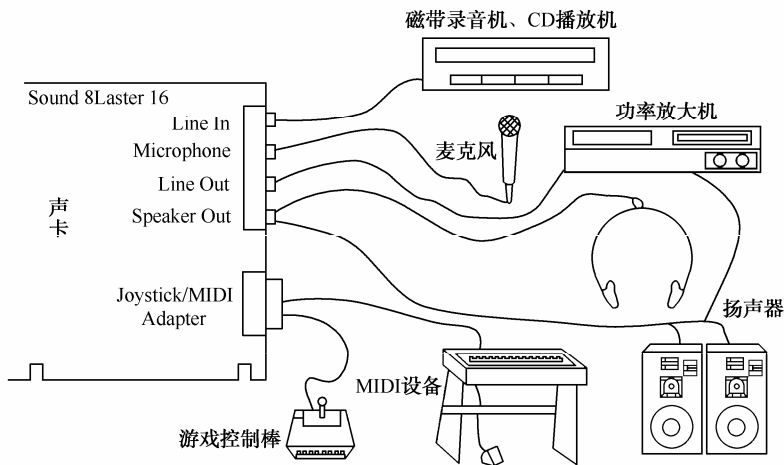


图 2.12 某 16 位声卡的接口与设备连接示意图

2.3 声音工具软件

声音工具（Audio Tools）用来录放、编辑和分析声音文件。声音工具使用相当普遍，但它们的功能相差很大。下面列出了比较常见的几种声音工具软件。

1. Windows操作系统自带的“录音机”

在操作系统中选择“开始”→“程序”→“附件”→“娱乐”→“录音机”之后就调出如图 2.13 所示的窗口。使用它可录音，做简单的声音编辑（如插入、删除等）（详见本书第二部分实验 1 数字声音）。

2. 买声卡时带的工具

如果你的计算机安装有声卡，一般来说都附带有声音工具。例如，声霸（Sound Blaster）卡带有几种声音工具，通常要由用户自己安装。其中，功能比较强的是 WaveStudio，它的用户界面如图 2.14 所示。



图 2.13 Windows 的录音机窗口

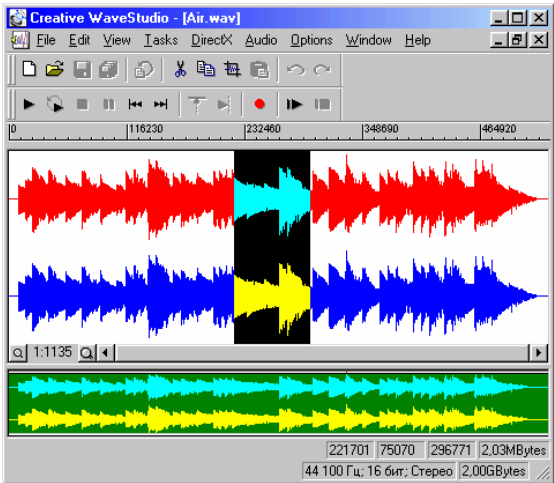


图 2.14 Creative WaveStudio 的用户界面

Creative WaveStudio 是一种方便易用的音频编辑工具，可播放、编辑和录音 8 位（磁带质量）、16 位（CD 质量）和 24 位的音频数据。使用各种特殊效果和编辑操作，如采用重复、

反转、回声、位移、剪切、复制和粘贴来增强你的音频数据或建立独特的声音，可支持 DirectX 音频插件。可以同时编辑多个音频文件，支持 RAW、MP3 和 Windows Media Audio（WMA）等格式的音频文件。

3. 网络下载工具

因特网上有许多站点提供试用的或者免费的声音工具。如图 2.15 所示为提供下载试用的 CoolEdit Pro（详见实验 T1.3）工具的用户界面。

CoolEdit Pro 是一个功能强大的音乐编辑软件，最高采样频率可以达到 192kHz，最高量化精度可以达到 32 位，支持 22 种音乐文件格式，能够高质量地完成录音、编辑、合成等多种任务，其完整的声音与音效的处理为用户提供了完整的音乐解决方案。只要拥有它和一台配备了声卡的计算机，就等于同时拥有了一台多轨数码录音机、一台音乐编辑机和一台专业合成器。

CoolEdit Pro 不仅适合于专业人员，也适合普通音乐爱好者。类似的工具还有 GoldWave，CakeWalk，Cubase 等。

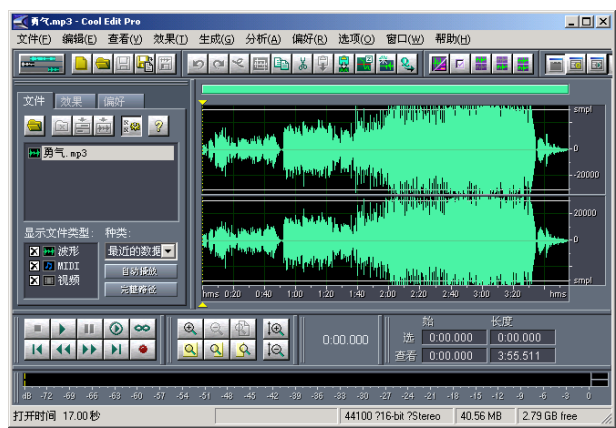


图 2.15 CoolEdit Pro 的用户界面

2.4 MIDI

如果你喜欢音乐，又爱玩计算机，那你肯定不会对 MIDI 这个词陌生。一直以来 MIDI 就是电脑音乐的代名词，MIDI 技术的多方优势和技术应用已经被广大音乐工作者和音乐爱好者所接受。在现代音乐制作上，MIDI 技术已达到不可取代的地位。在影视制作、游戏开发、广告配乐和计算机多媒体应用中都可以找到 MIDI 技术的影子。作为电脑音乐，MIDI 的优势在于其易修改性，可以表达传统乐器无法表现的特殊声效，即优良的音质和宽广的音域。精明的录音师可以方便地利用 MIDI 弦乐技术得到一支“永不出错的弦乐队”，好莱坞许多电影的配音已抛弃了传统乐器而完全围绕着 MIDI 进行了。随着计算机和因特网的迅速发展，MIDI 以其各种优势也开始在普通计算机爱好者中流行开来。很多人正是在听到真正优秀的 MIDI 音乐之后，才彻底改变了对电子音乐的看法。

2.4.1 MIDI规范与音乐合成

1. MIDI规范

MIDI 是 (Musical Instrument Digital Interface, 电子乐器数字接口) 的缩写, 是用于音乐合成器 (Music Synthesizers)、乐器 (Musical Instruments) 和计算机之间交换音乐信息的一种标准协议。从 20 世纪 80 年代初期开始, MIDI 已经逐步被音乐家和作曲家广泛接受和使用。MIDI 是乐器和计算机使用的标准语言, 是一套指令 (即命令的约定), 它指示乐器即 MIDI 设备要做什么, 怎么做, 如演奏音符、加大音量、生成音响效果等。MIDI 不是声音信号, 在 MIDI 电缆上传送的不是声音, 而是发给 MIDI 设备或其他装置, 让它产生声音或执行某个动作的指令。

MIDI 是由软件部分和硬件部分共同组成的系统规范, 这个规范定义了 MIDI 设备间信号传送时电缆硬件接口和协议。规范制定的目的是希望各厂商生产的乐器之间通过统一的 MIDI 交换信息及控制信号, 来完成音乐的合成。这样, 任何电子乐器, 只要有处理 MIDI 信息的处理器和适当的硬件接口都能变成 MIDI 装置。

MIDI 标准之所以受到欢迎, 主要是因为它有下列几个优点: 生成的文件比较小, 因为 MIDI 文件存储的是命令, 而不是声音波形; 容易编辑, 因为编辑命令比编辑声音波形要容易得多; 可以做背景音乐, 因为 MIDI 音乐可以和其他的媒体, 如数字电视、图形、动画、语音等一起播放, 这样可以加强演示效果。

2. 音乐合成

产生 MIDI 乐音的方法很多, 现在较为常用的方法有两种: 一种是 FM (Frequency Modulation) 合成法, 另一种是乐音样本合成法, 也称为波表 (Wave Table) 合成法。这两种方法目前主要用来生成音乐。

(1) FM 合成法

音乐合成器的先驱 Robert Moog 采用了模拟电子器件生成了复杂的乐音。20 世纪 80 年代初, 美国斯坦福大学 (Stanford University) 的一名叫 John Chowning 的研究生发明了一种产生乐音的新方法, 这种方法称为数字式频率调制合成法 (Digital Frequency Modulation Synthesis), 简称为 FM 合成器。他把几种乐音的波形用数字来表达, 并且用数字计算机而不是用模拟电子器件把它们组合起来, 通过数/模转换器 DAC (Digital to Analog Converter) 来生成乐音。斯坦福大学得到了发明专利, 并且把专利权授给 Yamaha 公司, 该公司把这种技术做在集成电路芯片里, 成了世界市场上的热门产品。FM 合成法的发明使合成音乐工业发生了一次革命。

FM 合成器生成乐音的基本原理如图 2.16 所示。它由 5 个基本模块组成: 数字载波器、调制器、声音包络发生器、数字运算器和 D/A 转换器。数字载波器用了 3 个参数: 音调 (Pitch)、音量 (Volume) 和波形 (Wave); 调制器用了 6 个参数: 频率 (Frequency)、调制深度 (Depth)、波形类型 (Type)、反馈量 (Feedback)、颤音 (Vibrato) 和音效 (Effect); 乐器声音除了有它自己的波形参数外, 还有它自己的比较典型的聲音包络线, 声音包络发生器用来调制声音的电平, 这个过程也称为幅度调制 (Amplitude Modulation), 并且作为数字式音量控制旋钮, 它的 4 个参数 Attack, Decay, Sustain, Release 写成 ADSR, 这条包络线也称为音量升降维

持静音包络线。

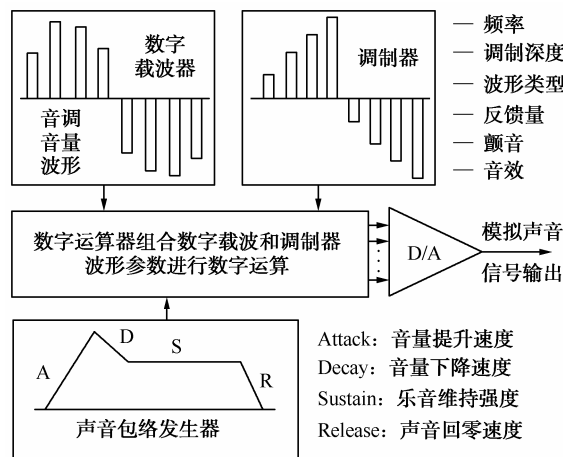


图 2.16 FM 合成器生成乐音的基本原理

在乐音合成器中，数字载波波形和调制波形有很多种，不同型号的 FM 合成器所选用的波形也不同。如图 2.17 所示是 Yamaha OPL-III 数字式 FM 合成器采用的波形。

各种不同乐音的产生是通过组合各种波形和各种波形参数，并采用各种不同的方法来实现的。用什么样的波形作为数字载波波形，用什么样的波形作为调制波形，用什么样的波形参数去组合才能产生所希望的乐音，这就是 FM 合成器的算法。

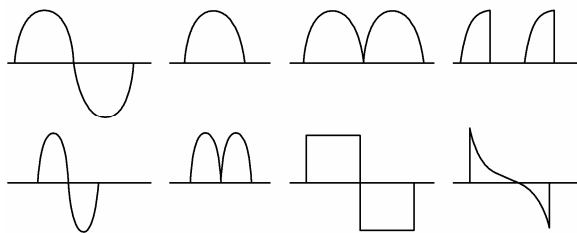


图 2.17 Yamaha OPL-III 数字式 FM 合成器采用的波形

通过改变图 2.16 中的参数，可以生成不同的乐音，例如：

- ① 改变数字载波频率可以改变乐音的音调，改变它的幅度可以改变它的音量；
- ② 改变波形的类型，如用正弦波、半正弦波或其他波形，会影响基本音调的完整性；
- ③ 快速改变调制波形的频率（即音调周期）可以改变颤音的特性；
- ④ 改变反馈量，就会改变正常的音调，产生刺耳的声音；
- ⑤ 选择的算法不同，载波器和调制器的相互作用也不同，生成的音色就不同。

在多媒体计算机中，图 2.16 中的 13 个声音参数和算法共 14 个控制参数以字节的形式存储在声音卡的 ROM 中。在播放某种乐音时，计算机就发送一个信号，这个信号被转换成 ROM 的地址，从该地址中取出的数据就是用于产生乐音的数据。FM 合成器利用这些数据产生的乐音是否真实，它的真实程度有多高，这就取决于可用的波形源的数目、算法和波形的类型。

(2) 波表合成法

使用 FM 合成法来产生各种逼真的乐音是相当困难的，有些乐音几乎不能产生。为了能真实地再现乐音，目前的声卡一般采用乐音样本合成法，即波表合成法。这种方法就是把真实乐器发出的声音以数字的形式记录下来，播放时改变播放速度，从而改变音调周期，生成各种音阶的音符，产生的声音质量比 FM 合成法产生的声音质量要高很多。乐音样本的采集相对比较直观。音乐家在真实乐器上演奏不同的音符，选择 44.1kHz 的采样频率、16 位的量化位数，这相当于 CD-DA 的质量，把不同音符的真实声音记录下来，这就完成了乐音样本的采集。

与 FM 合成法不同，波表合成法是采用真实的声音样本进行回放。乐音样本记录了各种真实乐器的波形采样，并保存在声卡上的 ROM 中。例如，创新的 Sound Blaster AWE32 是第一块广为流行的波表声卡。该卡采用了 EMU8000 波表处理芯片，提供 16 位 MIDI 通道和 32 位的复音效果。波表合成法的声音比 FM 合成法的声音更为丰富和真实，但由于需要额外的存储器存储音色库，因此成本也较高。而且音色库越大，所需的存储器就越多，相应的成本也就越高。乐音样本合成器的工作原理如图 2.18 所示。

波表合成法可以有软、硬之分，软波表原理跟硬波表一样，都是采用了真实的声音样本进行回放。只是硬波表的音色库是存放在声卡的 ROM 或 RAM 中的，而软波表的音色库则以文件的形式存放在硬盘里，需要时再通过 CPU 进行调用。由于软波表是通过 CPU 的实时运算来回放 MIDI 音效的，因此软波表对系统要求较高。

近几年随着 PCI 总线的流行而推出的 PCI 声卡，把硬波表和软波表的优点结合起来，提出了一种新的 MIDI 合成方案。其具体做法是：波表存储在硬盘上，使用时调入内存，但并非交给 CPU 处理，而是经 PCI 总线传回声卡，由声卡上的专用合成芯片处理。这一被称为“可下载样本”（Downloadable Sample）的合成技术（简称 DLS 技术），现已成为新一代 PCI 声卡的标准。硬盘上的样本库可选择 2MB、4MB 乃至 8MB 等不同的大小，音源与音质也可由用户选择，而且其内容可经常更新，使声卡的音频真正做到生动、灵活和多样。

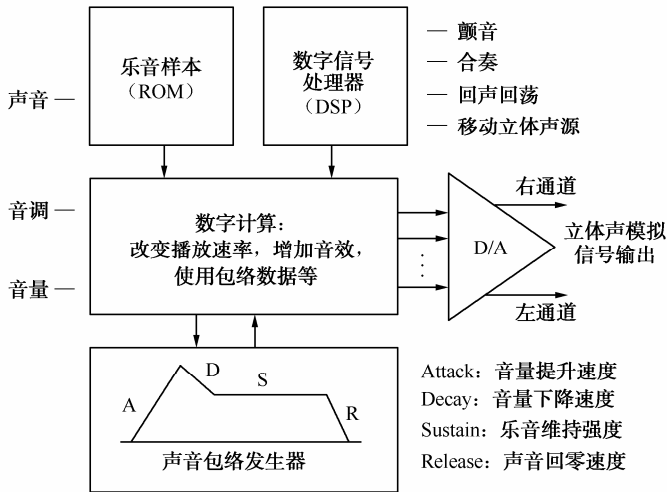


图 2.18 乐音样本合成器的工作原理

2.4.2 MIDI系统

MIDI 系统是指利用 MIDI 设备从事音乐创作时所利用的各种 MIDI 设备和其他电声设备的总和。一般来说，一个完整的 MIDI 系统包含 5 部分：输入设备、音序设备、发声设备、声音处理设备和还声设备，这些设备可以是独立的，也可以是集成的。MIDI 系统各部分的关系和连接如图 2.19 所示。

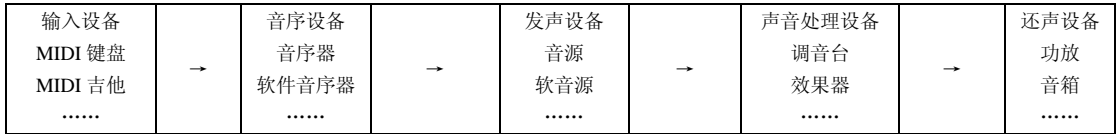


图 2.19 MIDI 系统各部分的关系和连接

1. 输入设备

为了符合人们原有的音乐习惯，人们制造了各种基于传统乐器形式上的 MIDI 乐器，如 MIDI 键盘（如图 2.20 所示）、MIDI 吹管、MIDI 吉他、MIDI 小提琴等，你可以按照演奏传统乐器的方式去演奏它们，然后你的演奏便通过 MIDI 接口被音序器接收并存储为音序内容。



图 2.20 具有 USB 接口的 MIDI 键盘

2. 音序设备

音序器是用来记录、编辑和播放 MIDI 文件的设备，它把一首曲子所需的音色、节奏、音符等按照一定的序列组织好让音源发声，它实际上是记录了音乐的一般要素，如拍子、音高、节奏、音符时值等。所有的音乐都必须由这些要素组成，音序器只不过是以数字的形式记录下它们，MIDI 文件的本质内容实际上就是音序内容。



图 2.21 Yamaha QY700 音序器

音序器分硬件和软件两种。最早的是硬件音序器，也成为编曲机。这类音序器的编辑和修改必须在它的面板上进行，使用是很不方便的。硬件音序器和音源连接以后就可以控制音源发声了。比较有名的硬件音序器有 Yamaha QY10，QY700（如图 2.21 所示）等。

软件音序器是安装在计算机中的多功能音乐创作、编辑

软件，如 Cakewalk, Encore, Cubase 等。软件音序器因为其显示界面的扩大（使用计算机显示器），功能的增强，特别是操作方便，获得了绝大多数音乐制作者的认可，大有淘汰硬件音序器之势。其实正是由于软件加入音乐制作的领域才使 MIDI 音乐和计算机联系起来，不然制作 MIDI 音乐完全用不着计算机。

3. 发声设备

由于音乐最终要以声音的形式表现出来，所以作为 MIDI 系统中产生声音的设备，音源和音色的数量、品种和质量都将对最终音乐作品的效果产生重要的影响。产生 MIDI 乐音的方法很多，现在用得较多的方法就是前面介绍的 FM（Frequency Modulation）合成法和波表（Wavetable）合成法。采用波表合成法的音源内部有很多不同音色的样本波形，如钢琴就有钢琴音色的真实样本，吉他就有吉他音色的真实样本（即波表）。但是音源只是一个资源库，因为它不知道在什么时候该用什么音色发怎样的声音。至于何时调用资源（使用哪种音色）不是音源的工作，这项任务是由 MIDI 制作的“心脏”——音序器来完成的。

音源也分硬件和软件两种。硬件音源是现在专业 MIDI 制作不可缺少的设备，因为它们可以提供比任何一块声卡上的波表都要好很多的音色，这些独立音源基本上是专业人士使用的，常见的型号有 Roland JV1080（如图 2.22 所示）和 Yamaha MU100R 等。



图 2.22 硬件音源 Roland JV1080

作为一般用户，可以不需要添置硬件音源设备，因为现在任何一块多媒体声卡上都有一个 128 种音色的 GM 音色库，即 MIDI 音源，只不过质量比专用音源设备差一些。但有些声卡（如 Sound Blaster Live）的音源质量较好，可达到准专业级。声卡自带的 MIDI 音源为 MIDI 的大众化做了很大的贡献，它使人们不需购买专门设备就可以进行 MIDI 的创作和演奏。

以前各生产商在设计 MIDI 发声工具（如声卡、音源器、合成器等）时并没有一个统一的乐器排序标准。例如，甲生产商把钢琴排序为第 1 号乐器声，而小提琴为第 40 号乐器声；但乙生产商却把钢琴排序为第 5 号乐器声，而小提琴则为第 60 号乐器声。那么，同一首乐曲如果使用不同生产商生产的 MIDI 发声工具来重播，便会用不同的乐器来演奏。原本是钢琴声的可能变成了吉他声，或原本是小提琴声的可能变成了手风琴声，这对用户来说十分不便。

基于这种情况，各生产商终于达成了共识，共同编制了一份“标准 MIDI 乐器排序表”（General MIDI Instrument Map, GM），如图 2.23 所示。GM 共收录了 16 类不同的乐器，每类各选 8 件，全部总共有 128 件乐器。在第 10 号 MIDI 通道（MIDI Channel 10）中收录了各种鼓和其他打击乐器。表中最后的一类（第 16 类）并非乐器声，而是一些声效（Sound Effects），如电话铃声、小鸟叫声、海浪声、直升机声、枪声等。

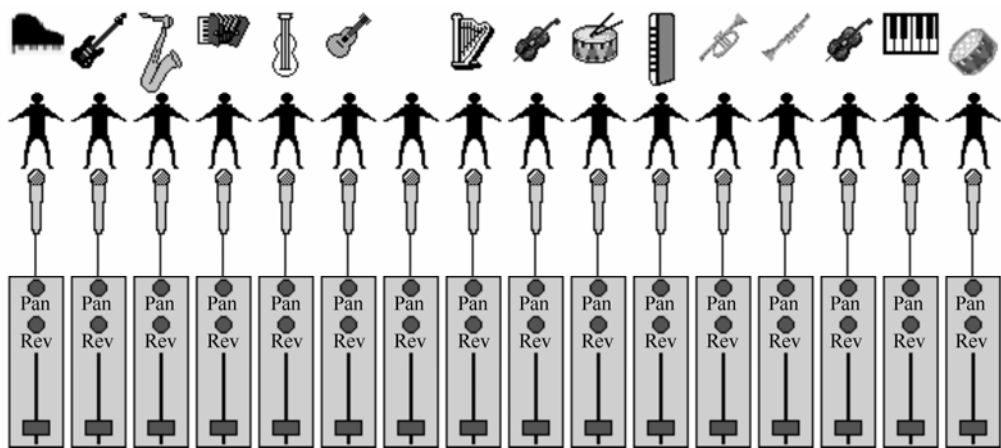


图 2.23 标准 MIDI 乐器排序表

4. 声音处理设备

通过调音台、效果器、均衡器等声音处理设备可对音频信号进行均衡、限幅、压缩、延时、混响等特技处理，也可以由软件进行处理，如 CoolEdit 就带有音效处理。

5. 还声设备

MIDI 系统还应该再配一个好一些的监听音箱，MIDI 音乐最后必须通过还声设备——音箱播放出来，能听到好的效果对于音乐制作是十分有利的。

基于 PC 的 MIDI 系统有多种构造方案，但基本都由以下几部分组成：多媒体个人计算机、音源、MIDI 键盘、音箱等。

MIDI 设备间的连接方式到目前一般有 3 种方式（即 3 种接口标准）：5 针 DIN 接口、USB 接口和串行接口。但目前家用计算机的声卡提供的 MIDI 接口是和声卡上的 GAME 摇杆接口共用一个 15 针的梯形接口，并提供 MIDI IN 和 MIDI OUT 端口。有专用的声卡连接线，如图 2.24 所示。



图 2.24 MIDI 设备的连接

计算机上安装的 MIDI 音序器软件、声效处理软件可完成音乐的创作、乐谱的打印，音响的幅度、节奏的快慢、各声部之间的协调及混响等。如图 2.25 所示是一个基于 PC 的 MIDI 系统。

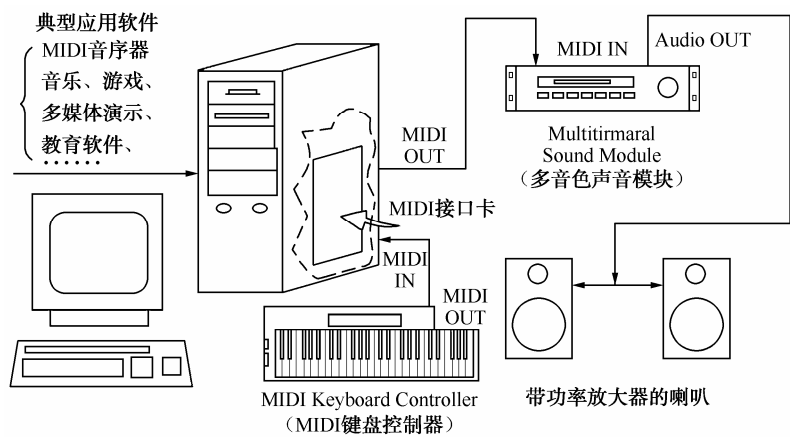


图 2.25 基于 PC 的 MIDI 系统

在这个系统中，PC 使用内置的 MIDI 接口卡，用来把 MIDI 数据发送到外部的多音色声音模块（音源），当然也可以不使用专业的硬件音源设备（如图 2.25 所示的多音色声音模块），而使用声卡自带的音源。像多媒体演示程序、教育软件或游戏等应用软件，它们把信息通过 PC 总线发送到 MIDI 接口卡。MIDI 接口卡把信息转换成 MIDI 消息（MIDI Messages），然后送到多音色声音模块同时播放出许多不同的乐音，如钢琴声、低音和鼓声。

使用安装在 PC 上的高级的 MIDI 音序器软件，用户可把 MIDI 键盘控制器（MIDI Keyboard Controller）连接到 MIDI 接口卡的中 IN 端口，从而具有音乐创作功能。计算机可通过音序器软件来采集 MIDI 键盘发出的一系列消息或指令。这一系列消息可记录在以“MID”为扩展名的 MIDI 文件中。在计算机上音序器可对 MIDI 文件进行编辑和修改。其工作原理如图 2.26 所示。

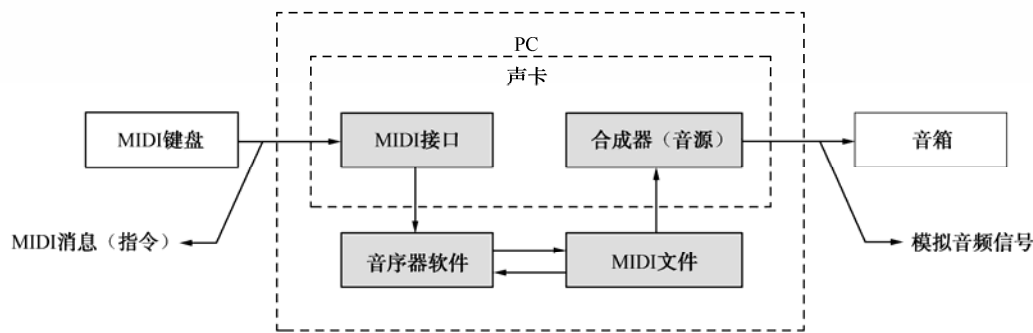


图 2.26 MIDI 的工作原理

2.4.3 MIDI 消息、文件和格式

MIDI 设备使用的一系列 MIDI 音符，可被认为是告诉音乐合成器（音源）如何播放一小段音乐的指令。因为 MIDI 数据是一套音乐符号的定义，而不是实际的音乐声音，因此 MIDI 文件的内容被称为 MIDI 消息（MIDI Messages）。一个 MIDI 消息由 1 个 8 位的状态字节并通

常跟着 2 个数据字节组成。在状态字节中，最高有效位设置成“1”，低 4 位用来表示这个 MIDI 消息是属于哪个通道的，4 位可表示 16 个可能的通道，其余 3 位的设置表示这个 MIDI 消息是什么类型的消息。MIDI 消息可分成通道消息（Channel Messages）和系统消息（System Messages）两大类，如图 2.27 所示。

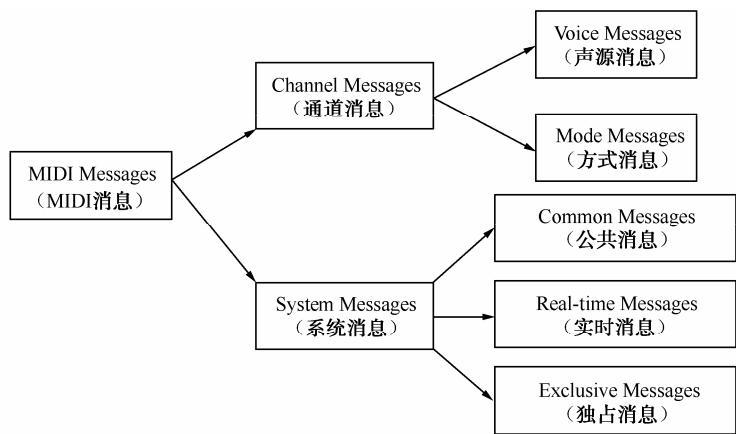


图 2.27 MIDI 消息

MIDI 通道消息可分成声源消息和方式消息，声源消息（Voice Messages）携带演奏数据，方式消息（Mode Messages）表示合成器响应 MIDI 数据的方式。MIDI 系统消息分成公共消息、实时消息和独占消息。公共消息（Common Messages）标志在系统中的所有接收器；实时消息（Real-time Messages）用于 MIDI 部件之间的同步；独占消息（Exclusive Messages）是厂商的标志代码。

国际 MIDI 协会（International MIDI Association）出版了标准 MIDI 文件（Standard MIDI Files）规范，定义了 3 种 MIDI 文件格式，扩展名为 MID。MIDI 音序器一般都能够处理多路 MIDI 数据流，或者叫“音轨”。标准 MIDI 文件格式使用格式 0（Format 0）在单独一个音轨中存储所有 MIDI 音序数据。格式 1（Format 1）文件则在多路音轨中保存 MIDI 数据，所有的音轨同时播放。格式 2（Format 2）文件也在多路音轨中保存 MIDI 数据，但所有音轨是独立播放的。大部分功能强大的 MIDI 音序器都可以读取格式 0 与格式 1 标准 MIDI 文件。格式 0 文件更小，因此节省存储空间，在传送方面格式 1 文件占用更少的带宽。不过格式 1 文件可以直接浏览及编辑，因此被更多音乐人所喜爱。

2.4.4 MIDI工具软件

既然 MIDI 音乐上的各种功能均可用 MIDI 消息来代表，那么我们只要编写出一些适当的软件来处理这些消息，便可实现利用计算机来处理音乐的梦想了。正因为如此，近几年来不断出现了大量不同功能的计算机音乐软件。有些是专为作曲及编曲而设计的，有些是专为制作和打印五线谱而设计的，有些是专为音乐教育而设计的，如 Piano、Music Lesson 等五花八门，目不暇接。有了这些软件的帮助，我们在学习音乐、作曲、编曲、制作和编印五线谱、制作唱片等各方面均发生了前所未有的变化。这与文字处理软件的出现，对出版、写作、印刷等方面所起的影响一样。

音序器是 MIDI 音乐创作的核心控制部件。在计算机系统中，音序器以纯软件的形式出

现，硬件资源完全由计算机和外部 MIDI 乐器提供。由于音序软件具有音序器所不具备的多种优点，如人-机交互界面、图形化的操作方式、更新升级方便等，因此音序软件已经逐步取代了音序器的功能，在专业音乐制作和音乐教学领域得到了广泛的应用。

Cakewalk 是 Twelve Tone Systems 公司设计的一个界面友好、功能强大的图形 MIDI 音序软件，也是世界范围内用户最多的音序软件。Cakewalk 提供了 MIDI 制作所需的全部功能。

国内使用最普遍的版本 Cakewalk Pro Audio 9.0X 的功能非常多，可提供 64 条音轨、256 条虚拟轨和 32 位的实时效果处理，并改进了文件管理系统，使装备了 Cakewalk Pro Audio 的计算机能够成为一台真正的多轨音频工作站。有多个控制图标可供用户在不同形式下观看并编辑 MIDI 音序、五线谱和事件表等，如图 2.28 所示。用鼠标可以控制 MIDI 的速度变化、调整力度或细调变音。可以在音序中嵌入波形声音文件。在 Cakewalk 中可使用 16 个 MIDI 接口设备，利用多媒体控制接口（MCI）可控制多媒体设备。

音乐工作站的未来发展方向是 MIDI、音频、音源（合成器）一体化制作。最先实现这个方式的是著名的 Cubase 软件。Cakewalk 公司奋起直追，于是推出了新一代的音乐工作站 Sonar。Cakewalk Pro Audio 9.0X 摇身变成了 Sonar，而现在 Sonar 已经升级到最新的 6.0 版本。Sonar 在 Cakewalk 的基础上，增加了针对软件合成器的全面支持，并且增强了音频功能，使之成为新一代全能型超级音乐工作站。



图 2.28 Cakewalk Pro Audio 9.0 音序软件主界面

2.5 其他语音技术

人的表达方式有多种，其中语音是最迅速、最常用和最自然的一种。让人与“机器”通过语言进行信息交流是科学家们多年来探索的领域。随着计算机技术的飞速发展，语音技术

将日益得到广泛的应用。

2.5.1 语音合成

20 世纪 90 年代中期，随着个人计算机的硬件和软件功能越来越强和现代语音技术的发展，以前在科幻电影中才能看到的会说话的计算机已经成为现实。而 TTS 技术正是计算机能够说话的关键技术之一。

语音合成（Speech Synthesis，或 Text To Speech，文语转换）简称 TTS，是将文本形式的信息转换成自然语音的一种技术，其最终目标是使计算机能够以清晰自然的声音，以各种各样的语言，甚至以各种各样的情绪来朗读任意的文本。也就是说，要使计算机具有像人一样，甚至比人更强的说话能力。因而它是一个十分复杂的问题，涉及语言学、韵律学、语音学、自然语言处理、信号处理、人工智能等诸多的学科。

语音合成就是一个将文本转化为语音输出的过程，这个过程的工作主要是将输入的文本按字或词分解为音素，并且对文本中的数字、货币单位、单词变形及标点等要特殊处理的符号进行分析，以及将音素生成数字音频，然后用扬声器播放出来或者存为声音文件以后用多媒体软件播放，这个过程如图 2.29 所示。

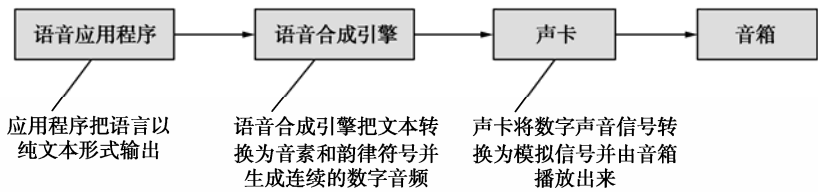


图 2.29 TTS 实现过程

当应用程序需要发声的时候就调用语音合成引擎（Speech Synthesis Engine）进行语言合成，将文本处理后通过音箱用近似于人的声音“读”出来。通常还可以通过改变对语音引擎的设置来改变“说话”的速度、声音频率（低沉或者尖锐）、声音大小，还能模拟口形、唇形和舌位的变化对声音的影响。

与一些用预先录制的声音文件实现发声的应用程序相比，TTS 的发声引擎只有几 MB 大小，不需要大量的声音文件支持，因此可以节省很大的存储空间，并且可以朗读预先未知的任何语句。现在已经有许多应用软件应用 TTS 技术实现语音功能。



图 2.30 Microsoft Excel 2003 中的文本到语音工具栏

【例 2.2】 Microsoft Office TTS 功能。

Microsoft Office 2003 中提供了 TTS 功能，可以对文本进行朗读。在 Microsoft Excel 2003 中，可选择菜单“工具”→“语音”→“显示语音到文本工具栏”，出现文本到语音工具栏，如图 2.30 所示。

Microsoft Word 2003 也可以对文本进行朗读。我们只要先选择的文本或者全部，然后选择发音操作，就可以对文本进行朗读。

另外，一些播音软件可以用来读小说或做校对工作，还可以朗读电子邮件，一些电子词典可以读出单词，还可以用于查询中心自动播放服务信息等。

文语转换在各种计算机相关领域中有着广泛的应用前景。当今，人与计算机之间进行交互的最常规手段是通过键盘输入信息，通过屏幕或打印机以视觉形式输出信息。这种方式不

同于人与人之间通过语音来交流信息的自然的交往方式，因而不仅极大地限制了广大普通用户使用计算机，而且在某些特定场合使用起来也很不方便。因此，构造一个以语音为媒介与计算机进行交互的系统，即智能计算机界面，是人们长久以来的梦想，也是科技人员孜孜以求的目标。显然，智能计算机界面包括两个相对独立的部分：“倾听”部分，即语音识别；“诉说”部分，即文语转换。随着这两方面技术的不断发展，将会从根本上改善人-机接口，从而使计算机以崭新的面貌进入人类生活，使计算机发挥出更大的作用。

2.5.2 语音识别技术

语音识别（Speech Recognition）就是研究让机器最终能听懂人类口述的自然语言的一门学科。语音识别以语音为研究对象，它是语音信号处理的一个重要研究方向，是模式识别的一个分支。语音识别比 TTS 更复杂，也较难以划分，涉及生理学、心理学、语言学、计算机科学及信号处理等诸多领域，甚至还涉及人的体态语言（如人在说话时的表情、手势等作为动作可帮助对方理解），其最终目标是实现人与机器进行自然语言通信。

语音识别系统可分为连续的与不连续的，如果语音识别系统是连续的，用户就能正常地与系统对话；如果不是连续的，用户就需要字字停顿。显然，连续语音识别（Continuous Speech Recognition）比不连续语音识别更受欢迎，但它需要更强的处理能力。一个典型语音识别系统的实现过程如图 2.31 所示。

预处理：包括语音信号采样，反混叠带通滤波，去除个体发音差异和设备、环境引起的噪声影响等，并涉及语音识别基元的选取和端点检测问题。

特征提取：用于提取语音中反映本质特征的声学参数，如平均能量、平均跨零率、共振峰等。

训练：在识别之前通过让讲话者多次重复语音，从原始语音样本中去除冗余信息，保留关键数据，再按照一定规则对数据加以聚类，形成模式库。

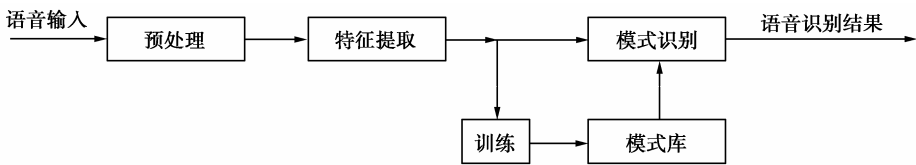


图 2.31 典型语音识别系统的实现过程

模式识别（匹配）：是整个语音识别系统的核心，它根据一定规则（如某种距离测度）及专家知识（如构词规则、语法规则、语义规则等）、计算输入特征与库存模式之间的相似度（如匹配距离、似然概率），判断出输入语音的语意信息。

进入 20 世纪 90 年代，随着多媒体时代的来临，迫切要求语音识别系统从实验室走向实用。许多发达国家如美国、日本、韩国，以及 IBM, Microsoft, Apple, AT&T, NTT 等著名公司都为语音识别系统的实用化开发研究投以巨资。

【例 2.3】 微软公司在 Microsoft Office 中的语音识别技术应用。

微软公司在 Microsoft Office 2003 中集成了语音识别技术，安装的时候可选择语音输入功能，如图 2.32 所示。



图 2.32 Microsoft Office 2003 中的语音输入

Microsoft Office 2003 安装了语音输入功能后，将麦克风连到声卡上的“MIC IN”，通过说话就可以输入文本到 Microsoft Word 2003 文档、Microsoft Excel 2003 单元格中。

在语音识别领域中，IBM 公司在世界上一直处于领先地位，不仅拥有 40 多年提供语音解决方案的经验，并且拥有 150 多个语音技术专利。IBM 公司领导了世界的语音识别技术，其语音识别产品在全球销售已达一百万套以上。IBM 公司在提供端对端的语音集成解决方案领域居全球领先地位。

1997 年 9 月，IBM 公司推出了 ViaVoice 中文连续语音识别软件系统（如图 2.33 所示），引起人们的极大关注。有人称 ViaVoice 是汉字输入的一次根本性革命，使计算机向人性化迈出了重要的一步，是中文信息处理技术发展的一个重要里程碑。ViaVoice 识别率可达 95% 以上，使用便利，特别适合于起草文稿、撰写文章和准备教案，是文职人员、作家和教育工作者的良好助手，还为广大上网的用户提供了轻松上网的功能，实现网上语音聊天的梦想。作为语音识别软件系列的产品，它可以使 PC、手提设备、汽车系统和自动客户服务系统之间的信息交流变得轻松快捷。作为第一个全功能的语音指令桌面程序，运行在 Windows 下的 ViaVoice 支持 Microsoft Office 2003，为不同要求的用户提供了精确的语音识别技术。

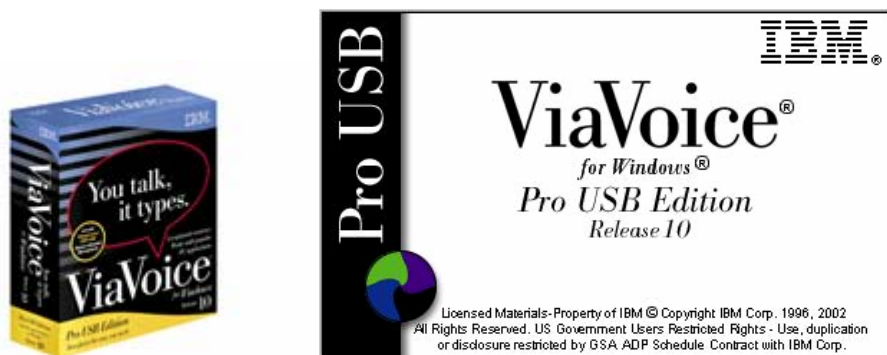


图 2.33 IBM ViaVoice 语音识别软件系统

语音识别和语音合成相结合，即构成一个完整的“人-机对话通信系统”。由于计算机的

迅速应用和普及，通过键盘、鼠标等手段的传统人-机对话模式已经不能适应人与机器之间的交互要求，人们非常希望能把人类之间快速、方便、直接和高效的通信方式——自然语音当做做人-机通信的媒介。当自然语言理解技术成熟后，语音识别技术将会得到更为广泛的应用，真正成为人们生活的一部分。

2.5.3 开发语音识别或语音合成方面的应用程序

实现 TTS 和 SR 的核心技术本身是非常复杂的，不是一般开发人员所能完成的，但是有了 Microsoft Speech SDK 这样的工具的帮助，只需要在应用程序中加入一些不太复杂的命令和操作，我们就可以创建具有 TTS 和 SR 功能的应用程序了。

Microsoft Speech SDK 是微软提供的软件开发包，提供的 Speech API（SAPI）主要包含两大方面：微软的连续语音识别引擎（MCSR）及微软的语音合成（TTS）引擎等。目前的 5.1 版本一共可以支持 3 种语言的识别（英语、汉语和日语）及 2 种语言的合成（英语和汉语）。SAPI 中还包括对于低层控制和高度适应性的直接语音管理、训练向导、事件、语法编译、资源、语音识别（SR）管理，以及 TTS 管理等强大的设计接口，其结构如图 2.34 所示。

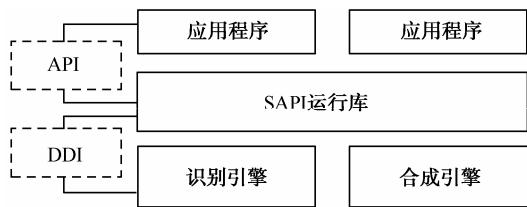


图 2.34 SAPI 结构

语音引擎则通过 DDI 层（设备驱动接口）和 SAPI（Speech API）进行交互，应用程序通过 API 层和 SAPI 通信。通过使用这些 API，用户可以快速开发在语音识别或语音合成方面的应用程序。金山词霸的单词朗读功能就用到了这些 API，而目前几乎所有的文本朗读工具都是用这个 SDK 开发的。

Microsoft Speech SDK 可以在微软的网站免费下载，目前的版本是 5.1，为了支持中文，还要把附加的语言包（LangPack）一起下载。

2.5.4 CD 的数据

在大多数播放软件的“打开文件类型”中，都可以看到.cda 格式，这就是 CD 音轨了。标准 CD 格式也就是 44.1kHz 的采样频率及 16 位量化位数，因为 CD 音轨可以说是近似无损的，因此它的声音基本上是忠于原声的，因此，如果你是一个音响发烧友的话，CD 是你的首选。CD 唱片可以在 CD 唱机中播放，也能用计算机里的各种播放软件来播放。

其实 CD 唱片（CD-DA）上的一首首歌曲，并非我们通常理解的一个个的文件。要知道，CD 唱片格式标准的确定之时，比计算机上用的 CD-ROM 格式（从 CD-ROM 开始光盘上才有文件系统）还要早，所以当初制定标准的时候当然不会考虑要让 CD-ROM 驱动器也能认出 CD 唱片。到后来，为了在计算机上更方便地使用 CD 音轨，就在计算机上规定：一个 CD 音轨为一个.cda 文件，如图 2.35 所示。这只是一个索引信息，并不是真正包含声音的信息，所以不论 CD 音乐的长短，在计算机上看到的.cda 文件大小都是 44 字节。

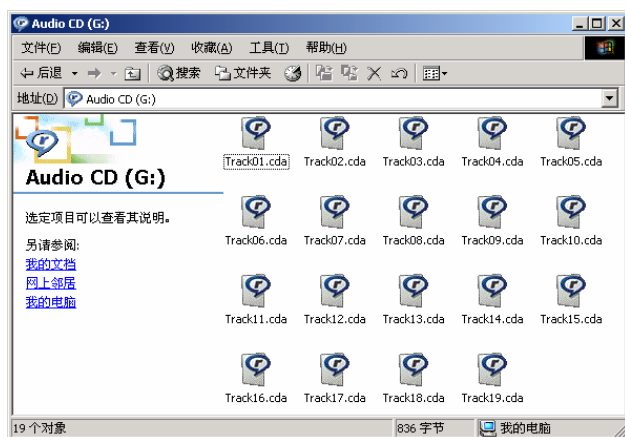


图 2.35 CD-DA 格式

注意，不能直接复制 CD 格式的.cda 文件到硬盘上播放，需要使用 Easy CD-DA Extractor (详见实验 T1.2)，EAC 这样的抓音轨软件可把 CD 格式转换成其他音频文件格式（如 MP3，WAV 等）。

习 题 2

1. 填空题

- (1) 人类能够接受的听觉带宽是从_____Hz 到_____kHz。
- (2) 声音信号数字化过程包括____、____、____3 个步骤。
- (3) 采用频率为 44.1kHz、量化精度为 16 位、持续时间为 1 分钟的 5.1 声道声音，未压缩时，数据量是_____MB。
- (4) MP3 是_____的简称，采用的是一种有损压缩方式。
- (5) 在 Intel 公司制定的_____音频规范中建议为了提高声音信号转换过程中的信噪比，减少电磁干扰，应该把数/模转换 (D/A) 和模/数转换 (A/D) 部分从主芯片中脱离出来，采用一个独立的处理单元来进行声音采样和编码。
- (6) 产生 MIDI 乐音的方法很多，现在用得较多的方法有两种：_____合成法和_____合成法。
- (7) 在 Windows 环境下，大部分的多媒体文件都是按一种文件结构来存放信息的，这个结构为“资源互换文件格式”，简称_____。
- (8) _____简称 TTS，是将文本形式的信息转换成自然语音的一种技术。

2. 简答题

- (1) 什么是语音？什么是全频带声音？
- (2) 什么是采样？简述 Nyquist 采样定理。
- (3) 什么是量化？若一个数字化声音的量化位数为 16，则能够表示的声音幅度等级是多少？
- (4) 常见的数字音频文件有哪些？

- (5) 声音数字化的主要参数有哪些？各有什么影响？
- (6) 简述声卡的主要功能、性能指标、工作原理。
- (7) 什么是 **MIDI**？什么是 **MIDI** 消息？
- (8) 一般来说，构成一个完整的 **MIDI** 系统，需要哪几部分设备？各起什么作用？
- (9) 合成音乐有哪两种方法？各有什么特点？
- (10) 谈谈语音合成和识别的应用领域，以及它们带来的变化。

第3章 数字图像

图像是人们非常容易接受的信息媒体，其最大的特点就是可以形象、生动和直观地表现大量的信息，具有文本和声音所不可比拟的优点。多少年来图像和计算机一直没有太多的联系，直到 20 世纪 70 年代，随着计算机图形学的发展，才可能通过计算机存储、处理和显示图像。图像数字化后形成数字图像，本章将介绍数字图像的基本概念与技术。

3.1 认识色彩

对于图像的设计与处理，认识色彩是创建完美图像的基础。在计算机上有一套特定的记录和处理色彩的技术。因此，要理解图像处理软件中所出现的各种有关色彩的术语，首先要具备基本的色彩理论知识。

3.1.1 色彩的来源

色彩是图像至关重要的组成部分，那么色彩是如何产生的呢？色彩源于光。1672 年牛顿用三棱镜将太阳光（白光）分解成红、橙、黄、绿、青、蓝、紫（品红）顺序排列渐变的彩带，这种现象称为色散（Dispersion），如图 3.1 所示。

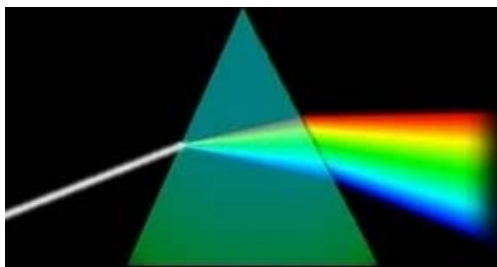


图 3.1 太阳光经过三棱镜后产生色散现象

色彩是通过光被我们所感知的，而光实际上是一种按波长辐射的电磁波。如图 3.2 所示，从电磁波谱及可见光光谱示意图可清楚地说明这一点。

不同的波段对应电磁波不同的波长，波长约 350~750nm 的光波能被人眼所接收，称为可见光。不同颜色的光实际上也对应于不同波长的光波。人眼根据光线的波长来感觉颜色，有色物体对光线具有选择性吸收的特性，即当光线照射到有色物体上时，入射光中被吸收的各种波长的色光是不等量的，有的被多吸收，有的被少吸收。

包含全部颜色光谱的光是白色，白光照射到有色物体上，其反射或透射的光线与入射光线相比，不仅亮度有所减弱，光谱成分也改变了，因而物体呈现出各种不同的颜色。而当没有光线时，人眼感觉的颜色是黑色的。

色彩的产生主要是因为人眼接收来自物体表面或内部对于光源的反射或透射，因此色彩和光有密切关系，同时还与被光照射的物体及观察者有关。

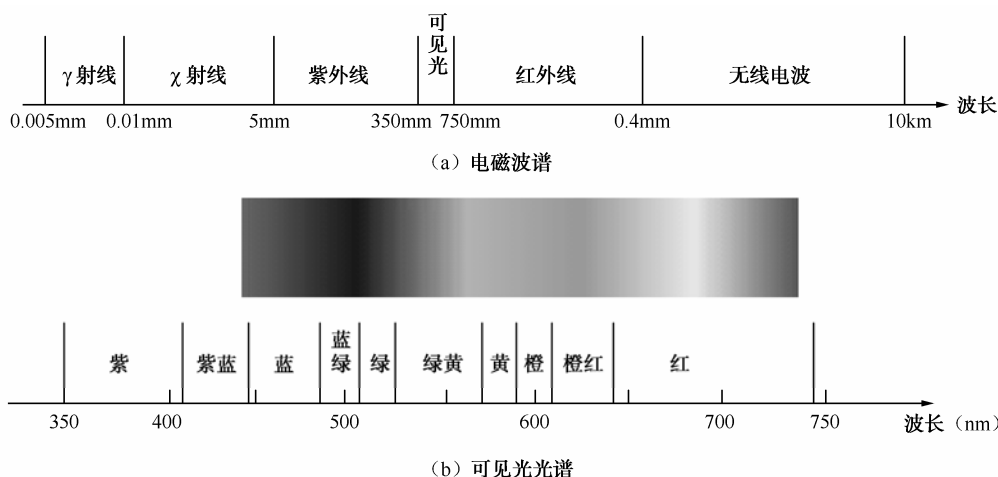


图 3.2 电磁波谱及可见光光谱示意图

3.1.2 色彩的三要素

从人的视觉系统看，色彩可用色调（Hue）、亮度（Lightness）和饱和度（Saturation）来描述。人眼看到的任一彩色光都是这 3 个特性的综合效果，这 3 个特性被称为色彩的三要素，其中，色调与光波的波长有直接关系，亮度和饱和度与光波的幅度有关。

1. 色调

光谱中各种颜色散发色彩的原始光辉，它们构成了色彩体系中的基本色调，如图 3.3 所示。在可见光光谱中，红、橙、黄、绿、青、蓝、紫每一种色调都有自己的波长和频率，人们给这些可以相互区分的颜色定出名称，当我们称呼到其中某一种颜色的名称时，就会有一个特定的色彩印象，这就是色调的概念。

物体的色调（Hue）与光波的波长有关，不同的波长反映不同的颜色感。色调是物体在日光照射下，所反射的光谱成分作用到人眼的综合效果，或者是当我们看到一种或多种波长的光时，所产生的彩色感觉。它反映颜色的种类，是决定颜色的基本特性。



图 3.3 光谱中的基本色调

2. 亮度

亮度（Lightness）是指光作用于人眼时所引起的明亮程度的感觉，是指色彩明暗深浅的程度。亮度有如下两种特性。

（1）同一物体因受光不同会产生亮度上的变化，如图 3.4（a）所示。照射的光越强，反射光也越强，看起来越亮。显然，如果彩色光的强度降到使人看不到了，在亮度标尺上它应与黑色对应。同样，如果其强度变得很大，那么亮度等级应与白色对应。

（2）强度相同的不同色光，亮度感不同，如图 3.4（b）所示。在纯正光谱中，黄色的亮度最高，显得最亮；其次是橙、绿；再其次是红、蓝；紫色亮度最低，显得最暗。

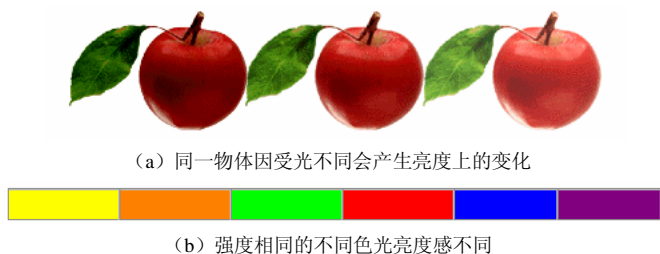


图 3.4 亮度的两种特性

3. 饱和度

饱和度 (Saturation) 是指颜色的纯度, 或者说是指颜色的深浅程度, 即掺入白光的程度。对于同一色调的彩色光, 饱和度越深颜色越鲜明, 或越纯。饱和度还和亮度有关, 同一色调越亮或越暗越不纯。如图 3.5 所示, 在红色光中加进白光之后, 相当于增加了光能, 因而变得更亮了, 但是它的饱和度却降低了, 红色被冲淡成粉红色; 若增加黑色光的成分, 相当于降低了光能, 因而变得更暗, 其饱和度也降低了。在某色调的彩色光中, 掺入别的彩色光, 会引起色调的变化, 而掺入白光时仅引起饱和度的变化。

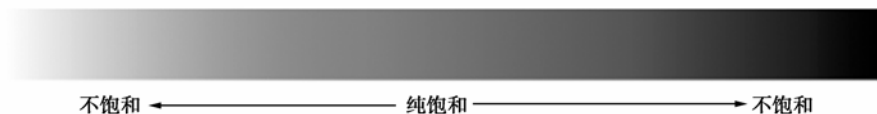


图 3.5 掺入白光对饱和度的影响

3.1.3 色彩模型

色彩模型是指彩色图像所使用的颜色描述方法, 也叫颜色模型。使用色彩模型的目的是尽可能多和有效地描述各种颜色, 以便需要时能方便地加以选择。各个应用领域一般使用不同的色彩模型, 如计算机显示时采用的是 RGB 模型, 彩色电视信号传输时采用 YUV 模型, 打印输出彩色图像时用 CMY 模型, 还有另外一些色彩模型的表示方法。下面将讨论几种常见的色彩模型。

1. RGB模型

自然界常见的各种颜色, 都可以由红 (R)、绿 (G)、蓝 (B) 3 种颜色光按不同比例相配而成, 如图 3.6 所示。同样, 绝大多数颜色光也可以分解成红、绿、蓝 3 种色彩。由于人眼对这 3 种色光最为敏感, RGB 3 种颜色相配所得到的彩色范围也最广, 所以一般都选这 3 种颜色作为基色, 这就是色度学的基本原理——三基色原理。

在多媒体计算机技术中, 因为计算机的彩色监视器的输入需要 RGB 3 个彩色分量, 通过 3 个分量的不同比例, 在显示屏幕上合成所需要的任意颜色, 所以不管多媒体系统中采用什么形式的色彩模型表示, 最后输出一定要转换成 RGB 彩色表示。

RGB 模型产生色彩的方式称为加色法, 因为没有光是全黑, 各色光加入后才产生色彩, 同时越加越高, 加到极限时成为白色。现在使用的彩色显示器和电视机都是利用这种三基色混合原理来显示彩色图像的, 而把彩色图片输入到计算机的彩色扫描仪则是利用它的逆过程。

扫描是把一幅彩色图片分解成 R, G, B 3 种基色, 每一种基色的数据代表特定颜色的强度, 当这 3 种基色的数据在计算机中重新混合时又显示出它原来的颜色。

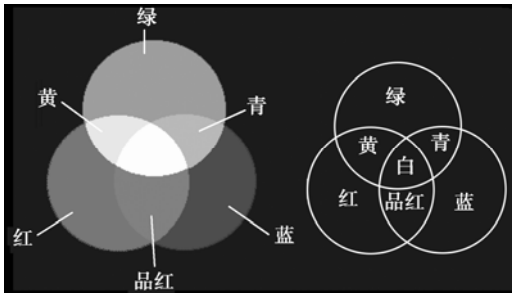


图 3.6 RGB 模型

【例 3.1】 用 RGB 模型设置 VB 标签控件 (Label1) 中字符串的颜色为红色, 如图 3.7 所示。



图 3.7 RGB 模型颜色显示示例

```
Private Sub Form_Load()  
    Label1.Caption = "RGB 颜色显示模型"  
    Label1.ForeColor = RGB(255, 0, 0) '红色  
End Sub
```

2. HSL模型

HSL (Hue, Saturation, Lightness) 模型是使用 H, S 和 L 3 个参数来生成颜色的。H 为颜色的色调, 改变它的数值可产生不同的颜色; S 为颜色的饱和度, 改变它可调节颜色的深浅程度; L 为颜色的亮度参量。HSL 模型如图 3.8 所示。

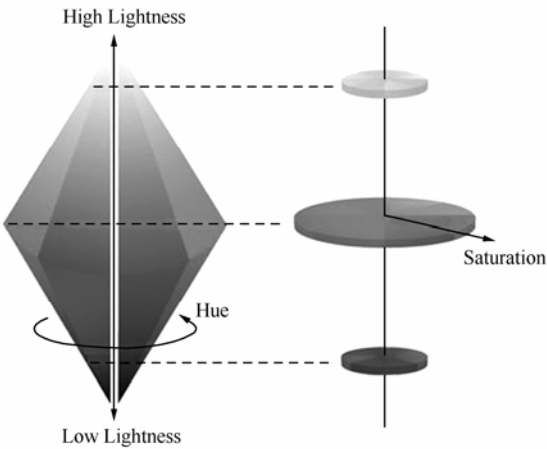


图 3.8 HSL 模型

由于人的视觉对亮度的敏感程度远强于对颜色浓淡的敏感程度，为了便于色彩处理和识别，人的视觉系统经常采用 HSL 色彩空间，它比 RGB 色彩空间更符合人的视觉特性。

【例 3.2】 Windows 操作系统画图软件中的 HSL 颜色模型。

用 Windows 系统画图软件画图，在“编辑颜色”对话框中显示了采用 HSL 和 RGB 色彩模型的对应关系。它们之间是一种线性关系，使得颜色的编辑十分直观、方便，有助于我们理解这两种色彩模型。这里，脸的颜色可以通过色调、饱和度和亮度设置为（228，204，158），然后填色即可，如图 3.9 所示。

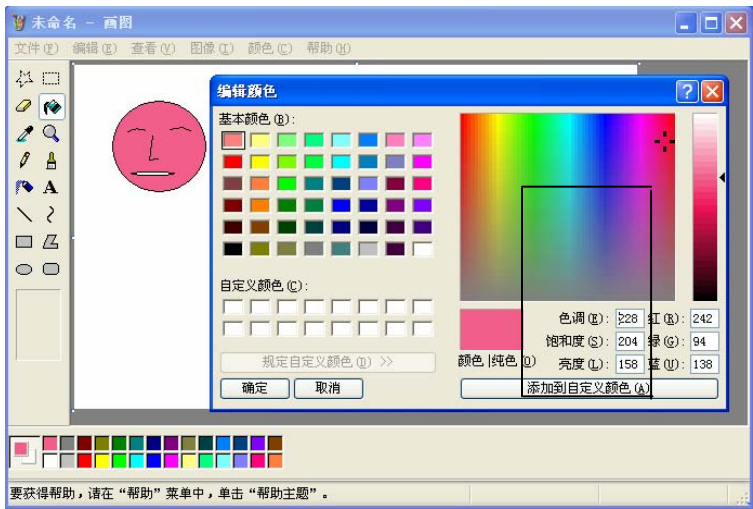


图 3.9 “编辑颜色”对话框

3. CMY模型

计算机屏幕显示彩色图像时采用的是 RGB 模型，而在打印时一般需要转换为 CMY 模型。CMY 模型（Cyan，Magenta，Yellow）是采用青、品红（洋红）、黄色 3 种基本颜色按一定比例合成颜色的方法，如图 3.10 所示。CMY 模型和 RGB 模型不同，色彩的产生不是直接来自于光线的色彩，而是由照射在颜料上反射回来的光线所产生的。颜料会吸收一部分光线（“减去”光），而未吸收的光线会反射出来，成为视觉判定颜色的依据，所以这种色彩的产生方式称为减色法。所有的颜料都加入后才能成为纯黑，当颜料减少时才开始出现色彩，颜料全部除去后才成为白色。

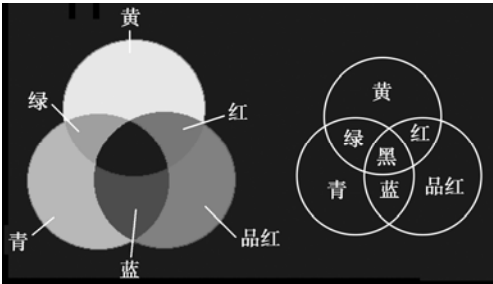


图 3.10 CMY 模型

虽然理论上利用 CMY 三原色混合可以制作出所需要的各种色彩,但实际上同量的 CMY 混合后并不能产生完善的黑色或灰色。因为所有打印油墨都会包含一些杂质,这 3 种油墨实际上产生一种土灰色,必须与黑色(K,由于字母 B 已经用来表示蓝色,因此黑色用字母 K 来表示)油墨混合才能产生真正的黑色。因此在印刷时必须加上一个黑色,这被称为 CMYK 模式。

四色印刷便是依据 CMYK 模式发展而来的。以常见的彩色印刷品为例,我们所看到的五颜六色的彩色印刷品,其实在印刷的过程中仅仅只用了四种颜色。在印刷之前先通过计算机或电子分色机将一件艺术品分解成四色,并打印成胶片。一般来说,一张真彩色图像的分色胶片是 4 张透明的灰度图,单独地看一张单色胶片时不会发现什么特别之处,但如果将这几张分色胶片分别以 C(青)、M(品红)、Y(黄)和 K(黑)4 种颜色叠印到一起再观察时,就产生了一张绚丽多姿的彩色照片。

【例 3.3】 彩色印刷品四色胶片。

某单位新年贺年福字贴图通过 Photoshop 设计,设计效果如图 3.11 所示。

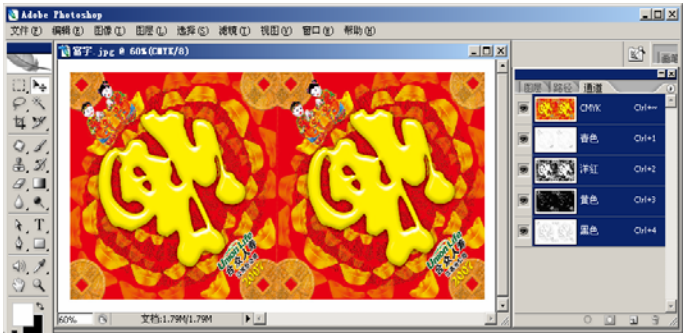


图 3.11 新年贺年福字贴图

通过通道标签,可以看到青色、洋红(品红)、黄色和黑色的情况。例如,选择“黑色”通道,新年贺年福字贴图中含有的黑色成分如图 3.12 所示。



图 3.12 “黑色”通道图

从图 3.12 中可以看出,图中黑色成分不太多。

在图设计完成后,按照 CMYK 模型,到印刷工厂发排,分别输出的四色胶片如图 3.13 所示。

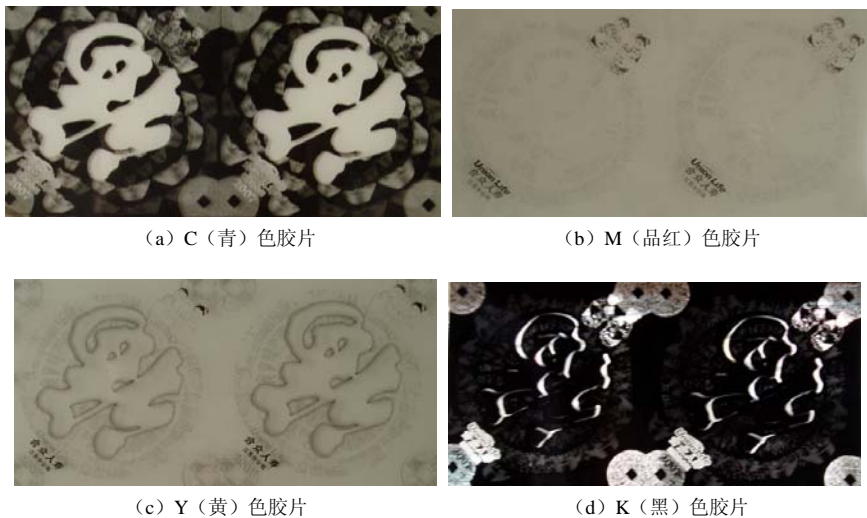


图 3.13 四色胶片

需要注意的是，K（黑）色胶片比“黑色”通道图要深得多，这是因为胶片深，透过的墨少，印刷到纸上的图上的黑色成分才可能少。所以，发排的图是通道图的反相。

分别用四色胶片控制 4 种颜色印刷，即可印刷出与设计图相同的图。

3.2 图像数据的获取与表示

计算机要对图像进行处理，首先必须获得图像信息并将其数字化。利用图像扫描仪、数码相机等最常用的图像输入设备对印刷品、照片或选定的景物进行拍摄，完成图像输入过程。下面将介绍数字图像数据的获取与表示的基本原理与相关知识。

1. 图像数据的获取

图像数据的获取是图像数字化的基础。图像获取的过程实质上是模拟信号的数字化过程。它的处理步骤大体分为 3 步，如图 3.14 所示。



图 3.14 图像数字化过程

(1) 采样。图像采样就是将二维空间上模拟的连续亮度（即灰度）或彩色信息，转化为一有限系列的离散数值来表示。由于图像是一种二维分布的信息，所以采样就是对图像在水平方向和垂直方向上进行等间隔的取样，每个采样点组成图像的基本单位，称为像素（pixel，Picture Element，简称为 pel），如图 3.15 所示。被分割的图像若水平方向上有 M 个采样点，垂直方向上有 N 个采样点，则一幅图像画面就被表示成 $M \times N$ 个像素构成的离散像素点的集合， $M \times N$ 表示图像的分辨率。

在进行采样时，采样点间隔的选取是一个重要的问题，它决定了采样后的图像是否能真实地反映原图像的程度，如图 3.16 所示。由于二维图像的采样是一维（声音）的推广，根据信号的采样定理——奈奎斯特（Nyquist）定理，要从采样样本精确地复原图像，图像采样的

频率必须大于等于原图像最高频率的两倍。

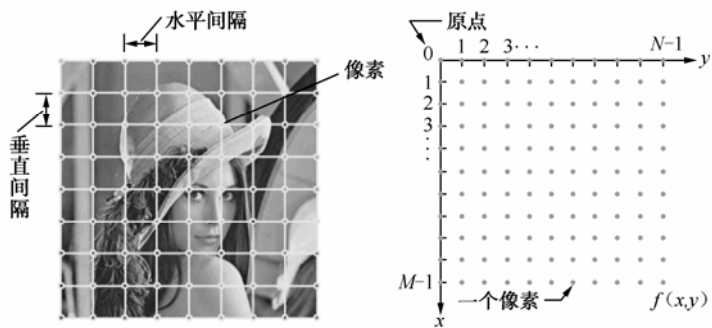
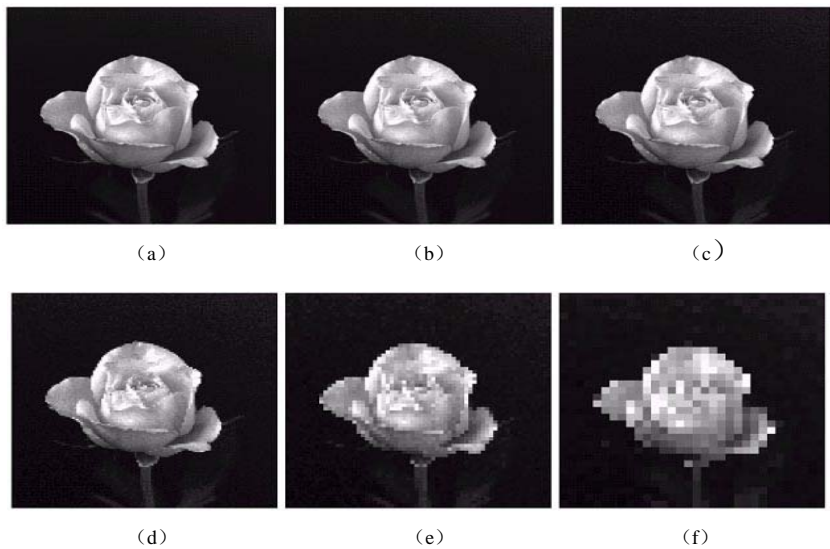


图 3.15 采样示意图



- (a) 1024×1024 图像
- (b) 对 (a) 图像采样 512×512 点复制相邻像素图像放大到 1024×1024 图像
- (c) 对 (a) 图像采样 256×256 点复制相邻像素图像放大到 1024×1024 图像
- (d) 对 (a) 图像采样 128×128 点复制相邻像素图像放大到 1024×1024 图像
- (e) 对 (a) 图像采样 64×64 点复制相邻像素图像放大到 1024×1024 图像
- (f) 对 (a) 图像采样 32×32 点复制相邻像素图像放大到 1024×1024 图像

图 3.16 图像的采样与图像的质量

(2) 量化。采样是对图像的空间坐标进行离散化处理，而量化是对每个离散点——像素的灰度或颜色样本进行数字化处理，把模拟量的亮度值用数字量来表示，如图 3.17 所示。

(3) 编码。数字化得到的图像数据量十分巨大，必须采用编码技术来压缩信息。从一定意义上讲，编码压缩技术是实现图像传输与存储的关键。



图 3.17 不同量化字长的灰度图像效果

2. 数字图像的表达

从数字图像的获取过程可以知道，一幅采样图像由 M （行）、 N （列）个采样点组成，每个采样点（像素）是组成图像的基本单位。黑白图像的像素只有 1 个亮度值，彩色图像的像素是矢量，它由多个彩色分量组成，一般有 3 个分量： R （红）、 G （绿）、 B （蓝），因此，采样图像在计算机中的表示方法是：单色图像用一个矩阵来表示；彩色图像用一组（一般是 3 个）矩阵来表示，矩阵的行数称为图像的垂直分辨率，列数称为图像的水平分辨率，矩阵中的元素是像素颜色分量的亮度值，使用整数表示。彩色图像的表达如图 3.18 所示。

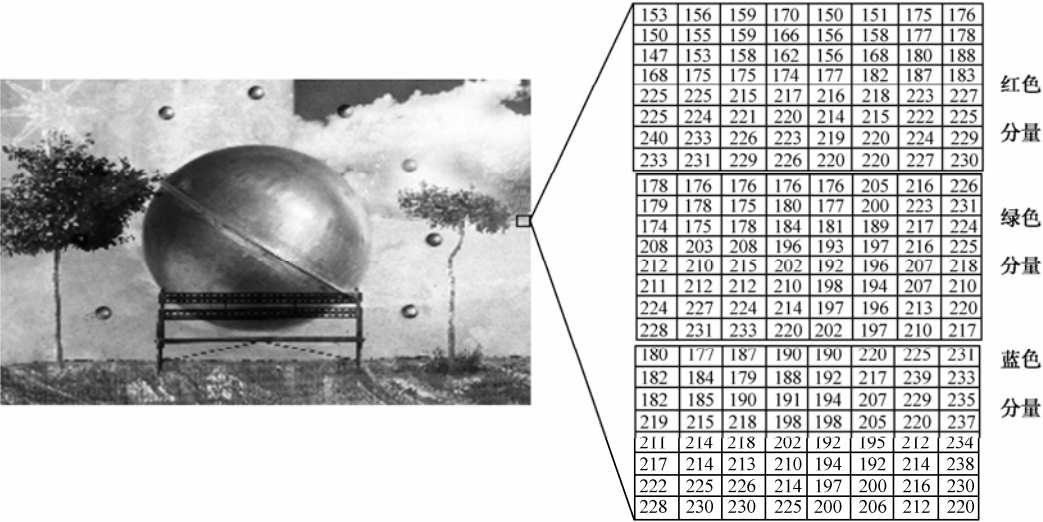


图 3.18 彩色图像的表达

3.3 图像的基本属性

在计算机中存储的每一幅数字图像，除了所有的像素数据外，描述一幅图像时还需要使用图像的属性。图像的属性包含分辨率、像素深度、图像大小、真/伪彩色、图像的种类等。

3.3.1 图像的分辨率

我们经常遇到的分辨率有两种：显示分辨率和图像分辨率。

1. 显示分辨率

显示分辨率是指显示屏上能够显示出的像素数目。例如，显示分辨率为 640×480 表示显示屏分成 480 行，每行显示 640 个像素，整个显示屏就含有 $640\times480=307\,200$ 个显像点。屏幕能够显示的像素越多，说明显示设备的分辨率越高，显示的图像质量也就越高。

显示分辨率确定了显示图像的区域大小，如果显示屏的分辨率为 640 像素 \times 480 像素，那么一幅 320 像素 \times 240 像素的图像只占显示屏的 1/4；相反，2400 像素 \times 3000 像素的图像在这个显示屏上就不能显示一个完整的画面。同样，不同分辨率的图像在相同的显示分辨率下显示的图像大小也不一样，如图 3.19 所示。



图 3.19 相同显示分辨率显示不同分辨率的图像（ $1024\times1024\sim32\times32$ ）

2. 图像分辨率

图像分辨率是指组成一幅图像的像素密度的度量方法。对同样大小的一幅图，如果组成该图的图像像素越多，则说明图像的分辨率越高，看起来就越逼真（包含了更多的图像细节）；相反，则图像显得越粗糙。如图 3.20 所示是不同图像分辨率的相同图像在相同大小显示区域中的显示情况。

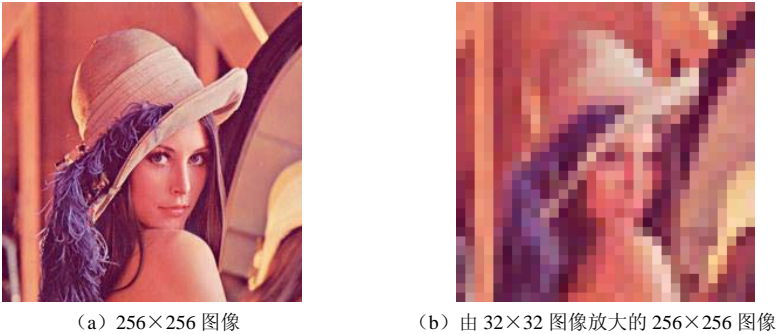


图 3.20 不同图像分辨率的相同图像在相同大小显示区域中的显示情况

在用扫描仪扫描彩色图像时，通常要指定图像的分辨率，用每英寸多少点即 dpi（dots per inch）表示。如果用 300dpi 来扫描一幅 8" \times 10" 的彩色图像，就得到一幅 2400 像素 \times 3000 像素的图像。分辨率越高，像素就越多。

图像分辨率与显示分辨率是两个不同的概念。图像分辨率确定组成一幅图像的像素数目，而显示分辨率确定显示图像的区域大小。

3.3.2 图像的像素深度

像素深度是指存储每个像素所用的位数，像素深度决定了彩色图像的每个像素可能有的颜色数，或者确定灰度图像的每个像素可能有的灰度等级数。深度为 1 位的图像只能有两种颜色（一般为黑色或白色）。灰度等级如果用 8 位来表示一个像素，即将纯黑和纯白间的层次等分为 256 级，就形成了 256 等级的灰度模式（如图 3.21 所示），就可以用来模拟黑白照片的图像效果。

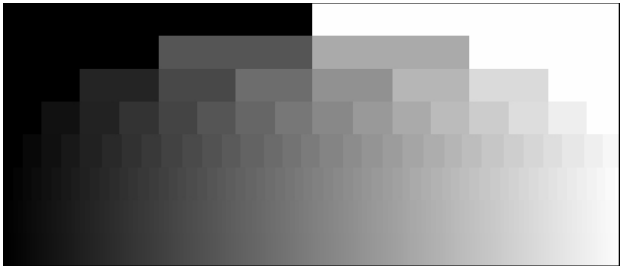


图 3.21 灰度等级参考图谱

对于彩色图像，每个像素用 R，G，B 3 个分量表示，若每个分量用 8 位，那么一个像素共用 24 位表示，每个像素可以是 $2^{24}=16\ 777\ 216$ 种颜色中的一种。从这个意义上，往往把像素深度说成是图像深度。表示一个像素的位数越多，它能表达的颜色数目就越多，而它的深度就越深。

虽然像素深度或图像深度可以很深，但各种 VGA 显示器的颜色深度却受到限制。例如，标准 VGA 显示器支持 4 位 16 种颜色的彩色图像，多媒体应用中推荐至少用 8 位 256 种颜色。由于设备的限制，加上人眼分辨率的限制，一般情况下，不一定要追求特别深的像素深度。此外，像素深度越深，所占用的存储空间越大。相反，如果像素深度太浅，也会影响图像的质量，图像看起来让人觉得很粗糙、很不自然。

在用二进制数表示彩色图像的像素时，除 R，G，B 分量用固定位数表示外，往往还增加 1 位或几位作为属性（Attribute）位。例如，RGB 5：5：5 表示一个像素时，用 2 字节共 16 位表示，其中 R，G，B 各占 5 位，剩下一位作为属性位。在这种情况下，像素深度为 16 位，而图像深度为 15 位。

属性位用来指定该像素应具有的性质。例如，在 CD-I 系统中，用 RGB 5：5：5 表示的像素共 16 位，其最高位（ b_{15} ）用于属性位，并把它称为透明（Transparency）位，记为 T。T 的含义可以这样来理解：假如显示屏上已经有一幅图存在，当这幅图或者这幅图的一部分要重叠在上面时，T 位就用来控制原图是否能看得见。例如，定义 $T=1$ ，原图完全看不见； $T=0$ ，原图能完全看见。

在用 32 位表示一个像素时，若 R，G，B 分别用 8 位表示，剩下的 8 位常称为 α 通道（Alpha Channel）位，Alpha 通道值也称为不透明度（Opacity）或覆盖率（Coverage Percentage），一个像素可用 4 个通道来表示，即（R，G，B， α ）。Alpha 通道借用不同的灰度值表示透明度的大小，8 位的 Alpha 通道值有 256 级透明度变化范围。每一级对应一个透明度，纯白（对

应 255 级) 为不透明, 纯黑 (对应 0 级) 为完全透明, 介于白黑之间的灰度值表示部分透明。

用这种办法定义一个像素的属性在实际中很有用。例如, 在一幅彩色图像上叠加文字说明, 而又不想让文字把图覆盖掉, 就可以用这种办法来定义像素, 而该像素显示的颜色又有人把它称为混合色 (Key Color)。在图像产品生产中, 也往往把数字电视图像和计算机生产的图像混合在一起, 这种技术称为视图混合 (Video Keying) 技术, 它也采用 Alpha 通道。

3.3.3 图像的大小

图像数字化后形成数字图像放在图像文件中, 图像文件的大小是指在磁盘上存储整幅图像所有像素的字节数 (Bytes), 可按下面的公式计算:

图像数据量=图像水平分辨率×图像垂直分辨率×像素深度/8

表 3.1 列出了若干不同参数的图像在压缩前的数据量。显然, 图像文件所需要的存储空间较大。在多媒体应用中, 一定要考虑图像的大小, 适当地掌握图像的分辨率和像素深度, 另外可采用数据压缩技术来解决部分问题。

表 3.1 几种常用格式的图像在压缩前的数据量

图像大小	8 位 (256 色)	16 位 (65 536 色)	24 位 (真彩色)
640×480	300KB	600KB	900KB
1024×768	768KB	1.5MB	2.25MB
1280×1024	1.25MB	2.5MB	3.75MB

3.3.4 图像的真彩色、伪彩色与直接色

在 RGB 色彩空间, 像素深度与色彩的映射关系主要有真彩色、伪彩色和直接色。搞清楚真彩色、伪彩色与直接色的含义, 对于编写图像显示程序、理解图像文件的存储格式有直接的指导意义。

1. 真彩色 (True Color)

真彩色是指将图像中的每个像素都分成 R, G, B 三个基色分量, 每个基色分量直接决定其基色的强度, 这样产生的色彩称为真彩色。例如, 图像深度为 24 位, 用 R : G : B = 8 : 8 : 8 来表示色彩, 则 R, G, B 各占用 8 位来表示各自基色分量的强度, 每个基色分量的强度等级为 $2^8=256$ 种。可生成 $2^{24}=16\,777\,216$ 种色彩, 这已超出了人眼所能分辨的颜色数量, 可以反映原图的真实色彩, 故称真彩色。

2. 伪彩色 (Pseudo Color)

伪彩色是指图像中每个像素的颜色不是由每个基色分量的数值直接决定的, 而是把像素值当做彩色查找表 (Color Look-Up Table, CLUT) 的表项入口地址, 去查找一个显示图像时使用的 R, G, B 强度值, 由查找出的 R, G, B 强度值产生的彩色称为伪彩色, 如图 3.22 所示。

彩色查找表 (CLUT) 是一个事先做好的表, 表项入口地址也称为索引号。例如, 16 种颜色的查找表, 0 号索引对应黑色, …… , 15 号索引对应白色。彩色图像本身的像素数值和彩色查找表的索引号有一个变换关系, 这个关系可以使用 Windows 操作系统定义的变换关系, 也可以使用你自己定义的变换关系。用查找得到的数值显示的彩色是真的, 但不是图像

本身真正的颜色，它没有完全反映原图的彩色。因此，同一幅图像，采用不同的颜色查找表显示可能会出现不同的色彩效果。

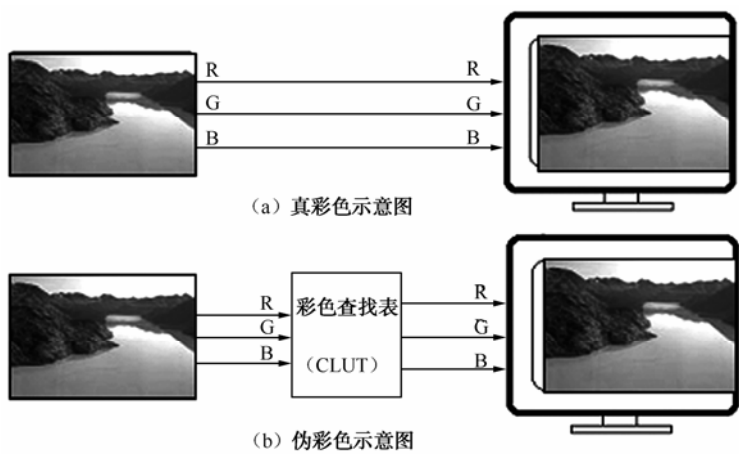


图 3.22 真彩色和伪彩色图像之间的差别

3. 直接色 (Direct Color)

每个像素值分成 R, G, B 分量，每个分量作为单独的索引值对它做变换。也就是通过相应的彩色变换表找出基色强度，由变换后得到的 R, G, B 强度值产生的彩色称为直接色。它的特点是对每个基色都进行变换。

用这种系统产生颜色与真彩色系统相比，相同之处是都采用 R, G, B 分量决定基色强度，不同之处是前者的基色强度直接由 R, G, B 决定，而后者的基色强度由 R, G, B 经变换后决定，因此这两种系统产生的颜色就有差别。试验结果表明，使用直接色在显示器上显示的彩色图像看起来真实、自然。

这种系统与伪彩色系统相比，相同之处是都采用查找表，不同之处是前者对 R, G, B 分量分别进行变换，后者是把整个像素当做查找表的索引值进行彩色变换。

3.3.5 图像的种类

1. 矢量图与位图

在计算机中，表达图像和计算机生成的图形、图像有两种常用的方法：一种是矢量图 (Vector Based Image) 法；另一种是位图 (Bit Mapped Image) 法。虽然这两种生成图的方法不同，但在显示器上显示的结果几乎没有什么差别。

矢量图是用一系列计算机指令来表示一幅图，如画点、画线、画曲线、画圆、画矩形等。这种方法实际上是用数学方法来描述一幅图，然后变成许多的数学表达式，再编程，用语言来表达。在计算显示图像时，也往往能看到画图的过程。绘制和显示这种图的软件通常称为绘图程序 (Draw Programs)。

矢量图有许多优点。例如，当需要管理每一小块图像时，矢量图法非常有效；目标图像移动、缩小、放大后不会失真，如图 3.23 (a) 所示，旋转、复制、属性的改变 (如线条变宽变细、颜色的改变) 也很容易做到。相同的或类似的图像可以把它们当做图像的构造块，并

把它们存到图库中，这样不仅可以加速画的生成，而且可以减小矢量图文件的大小。

然而，当图像变得很复杂时，计算机就要花费很长的时间去执行绘图指令。此外，对于一幅复杂的彩色照片（如一幅真实世界的彩照），恐怕就很难用数学来描述，因此不用矢量法表示，而采用位图法表示，但图像放大后会失真，如图 3.23（b）所示。

位图法与矢量图法很不相同。其实，位图已经在前面几节做了详细介绍，它是把一幅彩色图分成许多的像素，每个像素用若干个二进制位来指定该像素的颜色、亮度和属性。因此一幅图由许多描述每个像素的数据组成，这些数据通常称为图像数据，而这些数据作为一个文件来存储，这种文件又称为图像文件。若要画位图，或者编辑位图，则用类似于绘制矢量图的软件工具，这种软件称为画图程序（Paint Programs）。

位图的获取通常用扫描仪、摄像机、录像机、激光视盘与视频采集卡一类设备，通过这些设备把模拟的图像信号变成数字图像数据。



图 3.23 位图与矢量图放大后的比较

位图文件占据的存储器空间比较大。影响位图文件大小的因素主要有两个，即前面介绍的图像分辨率和像素深度。分辨率越高，组成一幅图的像素越多，则图像文件越大；像素深度越深，就是表达单个像素的颜色和亮度的位数越多，图像文件就越大。而矢量图文件的大小则主要取决于图的复杂程度。

矢量图与位图相比，显示位图文件比显示矢量图文件要快。矢量图侧重于“绘制”，去创造，而位图偏重于“获取”，去“复制”。矢量图和位图之间可以用软件进行转换，由矢量图转换成位图采用光栅化（Rasterizing）技术，这种转换也相对容易；由位图转换成矢量图用跟踪（Tracing）技术，这种技术在理论上说比较容易，但在实际中很难实现，对复杂的彩色图像尤其如此。

2. 灰度图与彩色图

灰度图（Gray-Scale Image）按照灰度等级的数目来划分。只有黑白两种颜色的图像称为单色图像（Monochrome Image），如图 3.24（a）所示为标准图像。图中的每个像素的像素值用 1 位存储，它的值只有“0”或者“1”，一幅 640×480 的单色图像需要占据 37.5KB（=640×480/8/1024）的存储空间。图 3.24（b）是一幅标准灰度图像。如果每个像素的像素值用一个

字节表示，灰度值等级数等于 256 级，每个像素可以是 0~255 之间的任何一个值，一幅 640×480 的灰度图像就需要占据 300KB（=640×480/1024）的存储空间。

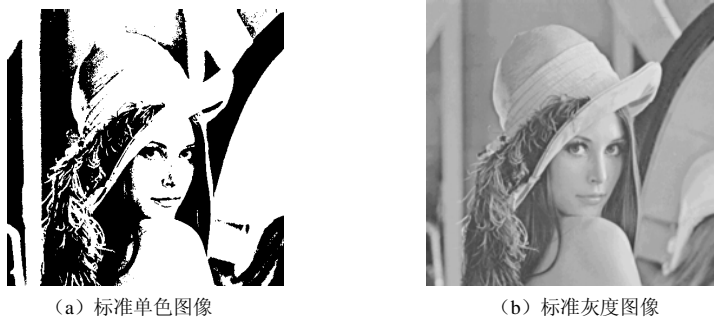


图 3.24 单色图像和灰度图像

彩色图像（Color Image）可按照颜色的数目来划分，例如，256 色图像和真彩色（ $2^{24}=16\,777\,216$ 种颜色）图像等。图 3.25（a）是一幅 256 色图像，彩色图像的每个像素的 R，G 和 B 值合用一字节来表示，一幅 640×480 彩色图像需要 300KB（=640×480/1024）的存储空间；图 3.25（b）是一幅真彩色图像，每个像素的 R，G，B 分量分别用一字节表示，一幅 640×480 的真彩色图像需要 900KB（=640×480×3/1024）的存储空间。



图 3.25 256 色图像和 24 位真彩色图像

许多 24 位彩色图像是用 32 位存储的，这个附加的 8 位叫做 α （Alpha）通道，它的值叫做 α （Alpha）值，多出的 8 位并没有提高颜色的种类，它用 8 位定义了一个透明度，也就是与其他图像之间的遮掩透明关系。

使用真彩色表示的图像需要很大的存储空间，用网络传输也很费时间。由于人的视觉系统的颜色分辨率不高，因此在没有必要使用真彩色的情况下应尽可能不用。

3.4 图像的文件格式

图像在存储媒体（如磁盘、光盘）中的存储格式，称为图像的文件格式。图像文件的格式有多种，如 BMP，GIF，JPEG，TIFF，PSD 等。

3.4.1 BMP 格式

BMP 是英文 Bitmap（位图）的简写，它是 Windows 操作系统中的标准图像文件格式，能够被多种 Windows 应用程序所支持。随着 Windows 操作系统的流行与多种 Windows 应用

程序的开发，BMP 位图格式理所当然地被广泛应用。这种格式的特点是包含的图像信息较丰富，几乎不进行压缩，但由此导致了它与生俱来的缺点——占用磁盘空间过大。所以，目前 BMP 在单机上比较流行。

BMP 文件由 4 个部分组成：位图文件头（Bitmap-File Header）、位图信息头（Bitmap-Information Header）、颜色表（Color Table）和位图数据，见表 3.2。

表 3.2 BMP 文件结构

位图文件的组成	结 构 名 称
位图文件头（Bitmap-File Header）	BITMAPFILEHEADER
位图信息头（Bitmap-Information Header）	BITMAPINFOHEADER
颜色表（Color Table）	RGBQUAD
位图数据	BYTE

1. 位图文件头

位图文件头数据结构含有 BMP 文件的类型、文件大小和位图起始位置等信息。其结构定义如下：

```
typedef struct tagBITMAPFILEHEADER
{
    WORD bfType;           // 位图文件的类型，必须为“BM”
    DWORD bfSize;          // 位图文件的大小，以字节为单位
    WORD bfReserved1;      // 位图文件保留字，必须为 0
    WORD bfReserved2;      // 位图文件保留字，必须为 0
    DWORD bfOffBits;       // 位图数据的起始位置，以相对于位图文件头的偏移
                          // 量表示，以字节为单位
} BITMAPFILEHEADER;      // 该结构占据 14 字节
```

2. 位图信息头

位图信息头数据用于说明位图的尺寸等信息。其结构定义如下：

```
typedef struct tagBITMAPINFOHEADER
{
    DWORD biSize;          // 本结构所占用的字节数
    LONG biWidth;          // 位图的宽度，以像素为单位
    LONG biHeight;         // 位图的高度，以像素为单位
    WORD biPlanes;         // 输出设备的位平面数，必须为 1
    WORD biBitCount         // 每个像素所需的位数，必须是 1（双色），4（16 色）
                          // 或 8（256 色）或 24（真彩色）之一（新的 BMP 格式支持 32 位色）
    DWORD biCompression;   // 位图压缩类型，必须是 0（不压缩），1（BI_RLE8 压
                          // 缩类型）或 2（BI_RLE4 压缩类型）之一
    DWORD biSizeImage;     // 位图的大小，以字节为单位
    LONG biXPelsPerMeter;  // 位图水平分辨率，每米像素数
    LONG biYPelsPerMeter;  // 位图垂直分辨率，每米像素数
```

```
        DWORD biClrUsed;           // 位图实际使用颜色表中的颜色数(为 0 则使用所有颜色
                                   // 表项)
        DWORD biClrImportant;      // 位图显示过程中重要的颜色数(为 0 则所有的颜色都是
                                   // 重要的)
    } BITMAPINFOHEADER;           // 该结构占据 40 字节
```

注意：对于 BMP 文件格式，在处理单色图像和真彩色图像的时候，无论图像数据多么庞大，都不对图像数据进行任何压缩处理。一般情况下，如果位图采用压缩格式，那么 16 色图像采用 RLE4 压缩算法，256 色图像采用 RLE8 压缩算法。

3. 颜色表

颜色表用于说明位图中的颜色，它有若干个表项，每一个表项都是一个 RGBQUAD 类型的结构，定义一种颜色。RGBQUAD 结构的定义如下：

```
typedef struct tagRGBQUAD
{
    BYTE rgbBlue;           // 蓝色的亮度 (值范围为 0~255)
    BYTE rgbGreen;          // 绿色的亮度 (值范围为 0~255)
    BYTE rgbRed;             // 红色的亮度 (值范围为 0~255)
    BYTE rgbReserved;        // 保留，必须为 0
} RGBQUAD;
```

颜色表中 RGBQUAD 结构数据的个数由 BITMAPINFOHEADER 中的 biBitCount 项来确定，当 biBitCount=1，4，8 时，分别有 2，16 和 256 个颜色表项，当 biBitCount=24 时，图像为真彩色，图像中每个像素的颜色用 3 字节表示，分别对应 R，G，B 值，图像文件没有颜色表项。位图信息头和颜色表组成位图信息，BITMAPINFO 结构定义如下：

```
typedef struct tagBITMAPINFO
{
    BITMAPINFOHEADER bmiHeader;    // 位图信息头
    RGBQUAD bmiColors[1];          // 颜色表
} BITMAPINFO;
```

注意：在 RGBQUAD 数据结构中，增加了一个保留字段 rgbReserved，它不代表任何颜色，必须取固定的值“0”，同时，在 RGBQUAD 结构中定义的颜色值中，红色、绿色和蓝色的排列顺序与一般真彩色图像文件的颜色数据排列顺序恰好相反，即若某个位图中的一个像素点的颜色的描述为“00，00，ff，00”，则表示该点为红色，而不是蓝色。

4. 位图数据

位图数据记录了位图的每一个像素值或该对应像素的颜色表的索引值，图像记录顺序是在扫描行内从左到右，在扫描行之间从下到上。这种格式我们又称为 Bottom_Up 位图，当然与之相对的还有 Up_Down 形式的位图，它的记录顺序是从上到下，对于这种形式的位图，也不存在压缩形式。位图的一个像素值所占的字节数：当 biBitCount=1 时，8 个像素占 1 字节；当 biBitCount=4 时，2 个像素占 1 字节；当 biBitCount=8 时，1 个像素占 1 字节；当 biBitCount=24 时，1 个像素占 3 字节，此时图像为真彩色图像。当图像不是真彩色时，图像

文件中包含颜色表，位图的数据表示对应像素点在颜色表中相应的索引值；当图像是真彩色时，每一个像素用 3 字节表示图像相应像素点的彩色值，每个字节分别对应 R，G，B 分量的值，这时候图像文件中没有颜色表。Windows 规定图像文件中一个扫描行所占的字节数必须是 4 的倍数（即以字为单位），不足的以 0 填充，图像文件中一个扫描行所占的字节数的计算方法：

$$\text{DataSizePerLine} = (\text{biWidth} * \text{biBitCount} + 31) / 32 * 4; \quad // \text{ 一个扫描行所占的字节数}$$

位图数据的大小按下式计算（不压缩情况下）：

$$\text{DataSize} = \text{DataSizePerLine} * \text{biHeight};$$

上述是 BMP 文件格式的说明，搞清楚了以上的结构，就可以正确地操作图像文件，对它进行读或写操作了。

3.4.2 GIF格式

GIF 是图形交换格式（Graphics Interchange Format）的英文缩写，是由 CompuServe 公司于 20 世纪 80 年代推出的一种高压缩比的彩色图像文件格式。1987 年开发的 GIF 文件格式的版本号是 GIF87a，1989 年进行了扩充，扩充后的版本号定义为 GIF89a。CompuServe 公司是一家著名的美国在线信息服务机构，针对当时网络传输带宽的限制，CompuServe 公司采用无损数据压缩方法中压缩效率较高的 LZW 算法，推出了 GIF 图像格式，主要用于图像文件的网络传输。GIF 文件不支持 24 位真彩色图像，最多只能存储 256 色的彩色图像或灰度图像，此外 GIF 允许用户为图像设置背景的透明属性。鉴于 GIF 图像文件的尺寸通常比其他图像文件（如 BMP）小好几倍，这种图像格式迅速得到了广泛的应用。考虑到网络传输中的实际情况，GIF 图像格式除了一般的逐行显示方式之外，还增加了交错显示方式，在图像传输过程中，用户可以先看到图像的大致轮廓，然后随着传输过程的继续而逐渐看清图像的细节部分，从而适应了用户的观赏心理，这种方式以后也被其他图像格式所采用，如 JPEG/JPG 等。最初的 GIF87a，GIF 只是用来存储单幅静止图像的，后来，又进一步发展成为 GIF89a，可以同时存储若干幅静止图像并进而形成连续的动画。目前 Internet 上大量采用的彩色动画文件多为这种格式的 GIF 文件。

1. GIF文件的结构

一个 GIF 文件的结构可分为文件头（File Header）、GIF 数据流（GIF Data Stream）和文件终结器（Trailer）3 个部分。文件头包含 GIF 文件署名（Signature）和版本号（Version）；GIF 数据流由控制标识符、图像块（Image Block）和其他的一些扩展块组成；文件终结器只有一个值为 0x3B 的字符（‘;’）表示文件结束。表 3.3 显示了一个 GIF 文件的组成结构。

表 3.3 GIF 文件结构

GIF 署名	文件头
版本号	
逻辑屏幕描述符	GIF 数据流

全局颜色表		GIF 数据流
.....		
图像描述符	图像块 (可重复 n 次)	
局部颜色表		
基于颜色表的图像数据		
.....		
GIF 结尾		文件终结器

(1) 文件头

GIF 署名 (Signature) 和版本号 (Version): GIF 署名用来确认一个文件是否是 GIF 格式的文件, 这一部分由 3 个字符 “GIF” 组成; 文件版本号也是由 3 个字节组成, 可以是 “87a” 或 “89a”。

(2) GIF 数据流部分 (GIF Data Stream)

① **逻辑屏幕标识符 (Logical Screen Descriptor):** 这一部分由 7 字节组成, 定义了 GIF 图像的大小 (Logical Screen Width & Height)、颜色深度 (Color Bits)、背景色 (Background Color Index), 以及有无全局颜色表 (Global Color Table) 和颜色表的索引数 (Index Count)。

② **全局颜色表 (Global Color Table):** GIF 格式中颜色表有全局颜色表和局部颜色表之分, 因为 GIF 格式允许一个文件中存储多个图像, 因此有这两种颜色表, 其中全局颜色表适于文件中的所有图像, 而局部颜色表只适用于某一个图像。全局颜色表必须紧跟在逻辑屏幕描述符后面, 每个颜色表索引条目由 3 字节组成, 按 R, G, B 的顺序排列。

③ **图像描述符 (Image Descriptor):** 一个 GIF 文件内可以包含多幅图像, 一幅图像结束之后紧接着是下一幅图像的描述符, 图像描述符以 0x2C (‘,’) 字符开始, 紧接着定义它的图像的性质, 包括图像相对于逻辑屏幕边界的偏移量、图像大小, 以及有无局部颜色表和颜色表大小, 由 10 字节组成。

④ **局部颜色表 (Local Color Table):** 如果上面的图像描述符中局部颜色表标志置位的话, 则需要在这里 (紧跟在图像描述符之后) 定义一个局部颜色表以供紧接着它的图像使用, 注意使用前应先保存原来的颜色表, 使用结束之后恢复原来保存的全局颜色表。如果一个 GIF 文件既没有提供全局颜色表, 也没有提供局部颜色表, 可以自己创建一个颜色表, 或使用系统的颜色表。

⑤ **基于颜色表的图像数据 (Table-Based Image Data):** 由两部分组成, 分别是 LZW 编码长度 (LZW Minimum Code Size) 和图像数据 (Image Data)。GIF 图像数据使用了 LZW 压缩算法, 大大减小了图像数据的大小。图像数据在压缩前有两种排列格式, 分别是连续的和交错的 (由图像描述符的交错标志控制)。连续方式按从左到右、从上到下的顺序排列图像的光栅数据; 交错图像按下面的方法处理光栅数据。

创建 4 个通道 (pass) 保存数据, 每个通道提取不同行的数据。

第 1 通道 (Pass 1) 提取从第 0 行开始每隔 8 行的数据。

第 2 通道 (Pass 2) 提取从第 4 行开始每隔 8 行的数据。

第 3 通道 (Pass 3) 提取从第 2 行开始每隔 4 行的数据。

第 4 通道 (Pass 4) 提取从第 1 行开始每隔 2 行的数据。

下面的例子（如图 3.26 所示）演示了提取交错图像数据的顺序。

行	通道 1	通道 2	通道 3	通道 4
0 -----	1			
1 -----				4
2 -----			3	
3 -----				4
4 -----		2		
5 -----				4
6 -----			3	
7 -----				4
8 -----	1			
9 -----				4
10 -----			3	
11 -----				4
12 -----		2		
13 -----				4
14 -----			3	
15 -----				4
16 -----	1			
17 -----				4
18 -----			3	
19 -----				4
20 -----		2		

图 3.26 GIF 提取交错图像数据的顺序

(3) 文件结尾部分

文件终结器 (Trailer)：这一部分只有一个值为 0x3B 的字节，标志一个 GIF 文件结束。

2. GIF 图像的显示

如果网络的速度过慢，或者图像过大，在浏览器中下载并显示图像时，不会立刻完成图像的下载，因此只能边下载边显示，这就出现了所谓的图像显示的中间过程。一般来说，在网页中显示图像时，图像是从上至下逐步显示出来的，如图 3.27 所示。这种显示方式的缺点是，在显示图像的过程中，用户在屏幕上无法看到图像的大致轮廓或整体结构，不能及时了解图像内容。

为了使用户在等待图像完全显示的过程中，尽快了解图像的内容，GIF 格式提供的交错显示方式，能帮助用户快速了解图像的整体结构。交错图在页面上逐步显示的时候，不是从上至下一段一段显示，而是以类似百叶窗的效果来显示。这样浏览者可以快速看到图像不同位置上的内容，尽管图像尚未显示完全，但是已经了解了大概。如图 3.28 所示为交错图像在页面上显示的过程。

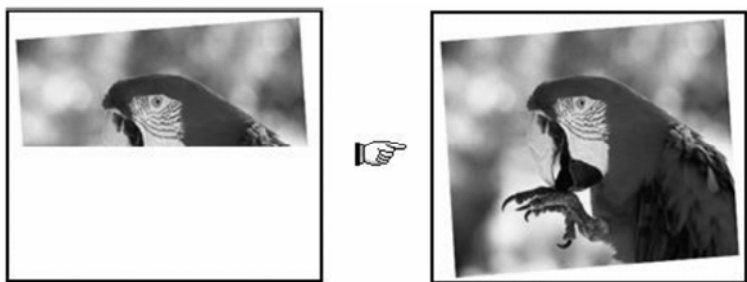


图 3.27 图像的正常显示过程

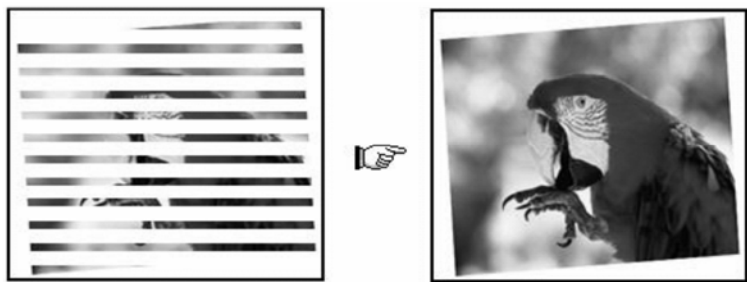


图 3.28 交错显示图像的过程

3.4.3 JPEG格式

JPEG (Joint Photographic Experts Group) 是国际标准化组织 (ISO) 和国际电报电话咨询委员会 (CCITT) 联合制定的静态图像压缩编码标准。JPEG 压缩技术十分先进，它用有损压缩方式去除冗余的图像数据，在获得极高的压缩率的同时能展现十分丰富、生动的图像。JPEG 格式压缩的主要是高频信息，对色彩的信息保留较好，适合应用于互联网，可减少图像的传输时间，可以支持 24 位真彩色，也普遍应用于需要连续色调的图像。

1. 支持多种压缩级别

JPEG 是一种很灵活的格式，具有调节图像质量的功能，允许用不同的压缩比例对文件进行压缩，支持多种压缩级别。压缩比率通常在 10：1 到 40：1 之间，压缩比越大，品质就越低；相反地，压缩比越小，品质就越好，如图 3.29 所示。



图 3.29 压缩比率对图像品质的影响

【例 3.4】 对 ACDSee 中文件格式转换工具压缩级别的控制。

ACDSee 是一个比较好的看图工具软件，新版本可以进行不同的图像文件格式转换。打开源图像文件，在转换成 JPEG 文件格式前显示“JPEG 选项”设置对话框，如图 3.30 所示，用户通过控制压缩比，可以控制生成的图像文件的大小和质量。



图 3.30 “JPEG 选项”设置对话框

2. 文件格式

JPEG 委员会对 JPEG 文件格式没有明确的定义，目前使用比较广泛的是 JPEG 文件交换格式（JPEG File Interchange Format, JFIF），此外还有 TIFF JPEG 等格式，但由于这种格式比较复杂，因此大多数应用程序都支持 JFIF 文件交换格式。

JPEG 文件由两部分构成：标记码和压缩数据。JFIF 文件格式直接使用 JPEG 标准为应用程序定义的许多标记，因此 JFIF 格式成了事实上的 JPEG 文件交换格式标准。标记码记录了 JPEG 图像的所有信息，在每个标记码之前可以有个数不限的填充字节 0xFF。下列是其中的 8 个标记：

- | | | |
|----------|-----------|-------------------|
| (1) SOI | 0xD8 | 图像开始 |
| (2) APP0 | 0xE0 | JFIF 应用数据块 |
| (3) APPn | 0xE1~0xEF | 其他的应用数据块（n, 1~15） |
| (4) DQT | 0xDB | 量化表 |
| (5) SOF0 | 0xC0 | 帧开始 |
| (6) DHT | 0xC4 | 哈夫曼（Huffman）表 |
| (7) SOS | 0xDA | 扫描线开始 |
| (8) EOI | 0xD9 | 图像结束 |

在 JPEG 文件存储格式中，每一个标记都是 JPEG 文件的一部分，因此 JPEG 文件实际上就是由上述的 8 个部分组成的。

3. 图像渐进显示

为了帮助用户在下载和显示图像时，快速了解图像的整体结构，GIF 格式提供了交错下载特性。而对于 JPEG 格式，它提供另外一种称为“渐进图”的格式，同样可以帮助用户及时了解图像整体信息。如果希望将 JPEG 图像以渐进图的格式保存，在 ACDSee 中，可以在将文件转换成 JPEG 格式时，在 JPEG 选项中选择“渐进”（见图 3.30），否则图像将以普通的 JPEG 格式保存。

如果图像是一幅渐进图，则它在页面上逐步出现的时候，会首先显示较模糊的整幅图像，然后逐渐变得清晰。这样，浏览者可以在下载过程中尽早了解图像的大致轮廓。如图 3.31 所示为一幅渐进图在页面上显示的过程。

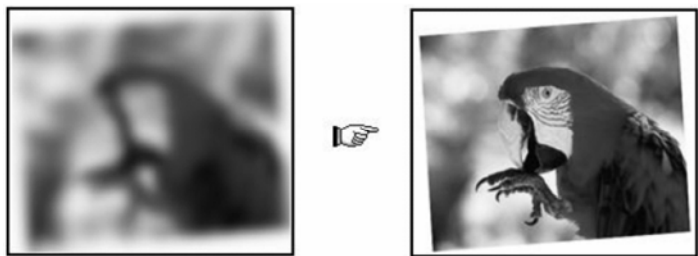


图 3.31 渐进图在页面上显示的过程

3.4.4 TIFF格式

TIFF（Tagged Image File Format）称为标记图像文件格式。它是 Aldus 和 Microsoft 公司为扫描仪和桌面出版系统研制开发的较为通用的图像文件格式。TIFF 的存储格式可以压缩也可以不压缩，压缩的方法也不止一种。TIFF 不依赖于操作环境，具有可移植性，PC 和苹果机同时支持 TIFF 格式。虽然 TIFF 格式的历史比其他的文件格式长一些，但现在仍是使用最广泛的行业标准位图文件格式，这主要是由于 TIFF 格式的规格经过多次改进。

最初设计 TIFF 的初衷就是要能够将扫描的图像在不同的平台上进行高质量的打印，所以 TIFF 格式比较适合作为高质量的保存原件的图像存储格式。TIFF 格式可以采用 RGB 模式或 CMYK 模式，因此 TIFF 图像在显示及打印两方面都能保持较高质量。

3.4.5 PSD格式

这是 Adobe 公司开发的图像处理软件 Photoshop 中自建的标准文件格式，它可以将所编辑的图像文件中的所有有关图层和通道的信息记录下来。所以，在编辑图像的过程中，通常将文件保存为 PSD 格式，以便于重新读取需要的信息。但是，PSD 格式的图像文件很少被其他软件和工具所支持。所以，在图像制作完成后，通常需要转换为一些比较通用的图像格式，以便于输出到其他软件中继续编辑。另外，在用 PSD 格式保存图像时，图像没有经过压缩，当图层较多时，会占很大的硬盘空间，图像制作完成后，除了保存为通用的格式以外，最好再存储一个 PSD 格式的文件备份，直到确认不需要在 Photoshop 中再次编辑该图像。由于 Photoshop 软件越来越广泛的应用，PSD 格式也逐渐流行起来。

3.5 图像系统组成

计算机图像系统应由输入、处理、输出 3 部分组成，本节主要介绍图像系统中常见的输入、输出设备。

3.5.1 显示器

显示器是计算机必不可少的一种图文输出设备，其作用是将数字信号转换为光信号，使文字与图像在屏幕上显示出来。没有显示器，用户便无法了解计算机的处理结果和所处的工

作状态，也无法进行操作。

当前显示器所采用的显示技术大体上可分为 CRT、液晶（LCD）等形式，它们有不同的技术特点。

1. CRT显示器

CRT（Cathode Ray Tube，阴极射线管）显示技术经过几十年的发展，已经是一个相对较为成熟的技术。CRT 显示器也是最常用的显示设备。CRT 显示器的原理并不复杂，它由显像管及相关的控制电路构成。显示器屏幕包含 3 种荧光粉，分别对应 3 种基本色（R，G，B）。显示器内的电子器件使用 3 种电子束，分别对应 3 种荧光粉。工作时，显像管后部的电子枪发出电子束，经过偏转磁场，轰击到荧光屏上某一点，使该点发光。通过变化电子束的强度，就能调整每种荧光粉发出的基本色度，合成不同的颜色。电子束从左到右、从上至下，逐行逐点轰击荧光屏，从而形成一帧图像，如图 3.32 所示。

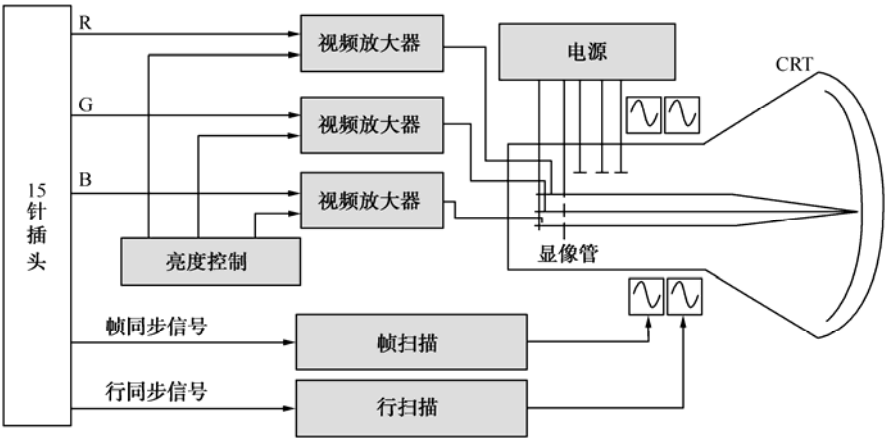


图 3.32 CRT 显示器的原理

台式 PC 一般使用 CRT 显示器，其中首选的是平面显像管构成的显示器，其屏幕在水平和垂直方向几乎是笔直的，因而显示器的失真和反光被减少到最低限度，视觉效果好，从任何角度看画面均无扭曲现象发生，显示效果极佳。

这些纯平显像管可以被分为荫罩式（Shadow-Mask）显像管和荫栅式（Aperture-Grille）显像管。

（1）荫罩式显像管显示器

荫罩式彩色显像管很早以来就被广泛应用。最早的荫罩式显像管是三枪三束式的。1973 年自会聚式显像管研制成功。到 1994 年前后，开始出现了商业化生产的“平面直角”荫罩式彩色显像管。到 1998 年年底，纯平荫罩式彩色显像管开始上市。目前纯平荫罩式彩色显像管市场上占有率最大的是 SAMSUNG，LG 等厂商。

荫罩是一块刻有成千上万个孔的薄钢板。荫罩孔的作用在于保证 3 个电子共同穿过同一个荫罩孔，以激发荧光粉，使之发出红、绿、蓝三色光。目前荫罩式纯平显像管所用的荫罩主要是孔状荫罩（如图 3.33（a）所示）和沟槽状荫罩（如图 3.33（b）所示）。前者以 SAMSUNG 为代表，后者以 LG 为代表。

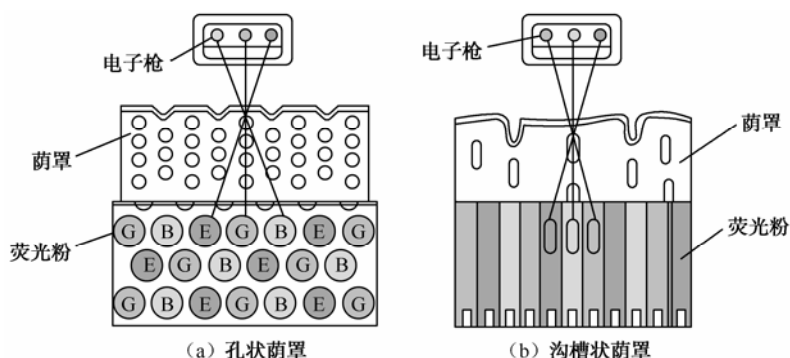


图 3.33 两种不同的荫罩

SAMSUNG 的纯平显像管 DynaFlat 采用孔状荫罩这种传统的设计的一个好处是有效地控制了成本，但缺点是画面不够精细，屏幕上的颗粒感比较明显，而且由于点状荫罩的固有缺点，造成电子透过率在 50% 左右，难以进一步提高亮度和对比度。而 LG 的“未来窗”则采用了沟槽状的荫罩，画面精细程度有所提高，但还属于荫罩式，电子透过率在 70% 左右，亮度和对比度进一步增加。但是，由于荫罩式显像管本身结构的限制，电子透过率已经难以进一步提高，所以目前在主流纯平市场上的这种显像管一般只用于中低档显示器中。

当一束电子穿过一个荫罩孔，必定准确地投射在一组 R, G, B 荧光点所组成的像素点上，荧光屏上的像素点排列的紧密程度与荫罩孔排列的紧密程度是相关的，所以也有厂商用荫罩孔距来标称点距，如图 3.34 所示。电视机用的 CRT 的点距一般为 0.78mm，而标准计算机显示器的点距为 0.28mm。点距越小的显示器屏幕越清晰，就需要更小更精细的荧光点，对于显像管的聚焦性能要求就越高。这也就是为什么同样尺寸的计算机显示器比电视机的价格贵得多的原因。

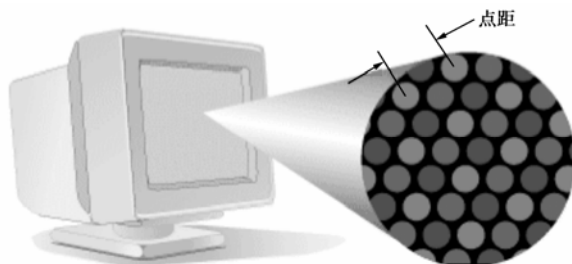


图 3.34 CRT 显示器的点距

(2) 荫栅式显像管显示器

荫栅式显像管在 20 世纪 80 年代开始加速发展。1982 年，SONY 首先把“单枪三束”的“特丽珑”（TRINITRON）技术引入显示器领域，实现了显示技术的重大突破，成为专业显示器的首选。SONY 出于战略性考虑，将技术转让给了三菱，三菱在 SONY 的技术上发展出“三枪三束”的“钻石珑”（DIAMONOTRON）技术。目前，SONY 的 FD TRINITRON 和三菱的 DIAMONDTRON NF 显像管牢牢地占据着高端纯平显像管市场。在国内，SONY 也通过战略合作伙伴“雅美达”（ARTMEDIA）向中端显示器市场进军。

荫栅式显像管的技术特点就是将荧光粉安排成跨越整个显示器屏幕的竖条状，将荫罩改为条状荫栅，如图 3.35 所示。这些条状荫栅由固定在一个拉力极大的铁框中的互相平行的铁线阵列组成。这样设计的好处是铁线是互相平行的，在垂直方向上没有任何东西阻挡电子通过，增加了电子的透通过率，使电子透过率达到 95% 以上，远远超过了荫罩结构的显像管，亮度和色彩饱和度更好，画面细腻动人，没有颗粒感，这也是雅美达显示器图像显示效果出色的重要原因。由于吸收电子少，长时间使用荫栅也不会由于受电子束冲击产生热量引起膨胀或变形，避免了出现颜色突变和色彩减低的情况。由于自身的技术特点，普通的 FD TRINITRON 就可以达到 0.24mm 的精细栅距（栅距概念见下文），而 21 英寸的产品更可以达到超精细的 0.22mm，画面的精细程度进一步提高。因此，目前在专业显示器中，绝大多数都采用了荫栅式显像管。而技术比较成熟可靠的 TRINITRON 在其中占大部分，据统计有 70% 以上的专业显示器采用了 SONY 的显像管。在采用 TRINITRON 显像管的厂商中，雅美达是其中的佼佼者。

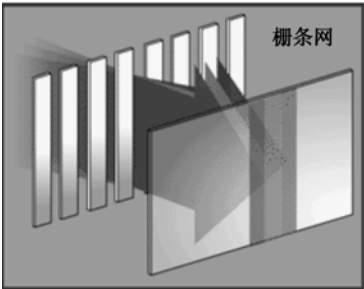


图 3.35 荫栅式显像管示意图

荫栅式显像管的缺点就是为了保持铁线的平行，需要在横向设置阻尼线。FD TRINITRON 的阻尼线比柱面 FD TRINITRON 的更加细小，更难发现它的存在，对用户的使用一般不会造成什么影响。另外，由于制造工艺要求比较高，它的成本要比荫罩式纯平显像管高一些。但是就目前来看，荫栅式显像管，特别是 FD TRINITRON 显像管，无疑是中高端用户的最佳选择。

荫栅式显像管不存在点距的概念。这种显像管的彩色元素是由红、绿、蓝 3 色的竖向条纹构成的，没有像素点，因此我们引入了“栅距”（如图 3.36 所示）这个概念，所谓的栅距也就是栅条之间的距离。

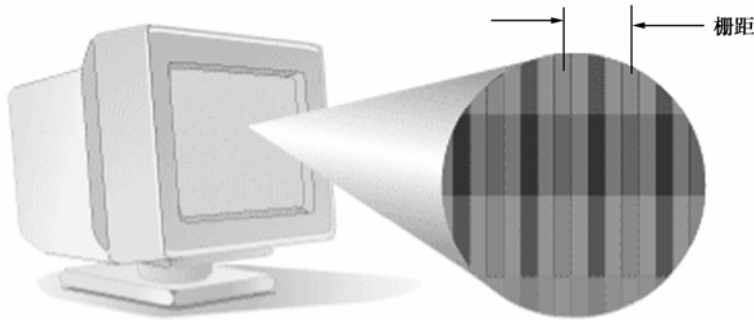


图 3.36 CRT 显示器的栅距

2. 液晶显示器

目前，液晶显示（Liquid Crystal Display，LCD）技术呈现出强劲的发展势头，大有取代 CRT 显示器的趋势。

液晶于 1888 年由奥地利植物学者 Reinitzer 发现，是一种介于固体与液体之间，具有规则分子排列的有机化合物。在不同电流电场作用下，液晶分子会做规则旋转，产生透光度的差别，在电源开/关上产生明暗的区别，在彩色 LCD 面板中，每个像素都是由 3 个液晶单元格构成的，其中每一个单元格前面都分别有红色、绿色和蓝色的过滤器。这样，通过不同单

元格的光线就可以在屏幕上显示出不同的颜色。依此原理控制每个像素，便可构成所需的图像。下面重点介绍 TFT 型液晶显示器的原理。

如图 3.37 所示，背光源发光，也就是荧光灯管投射出光源，这些光源会先经过一个偏光板后再经过液晶分子，分子的排列方式可改变穿透液晶的光线角度，然后这些光线接下来还必须经过前方的彩色滤光膜与另一块偏光板导出。位于底层的薄膜式电晶体，可通过改变液晶的电压值来控制最后出现的光线强度，并进而能够在液晶面板上组合出有不同深浅的颜色。

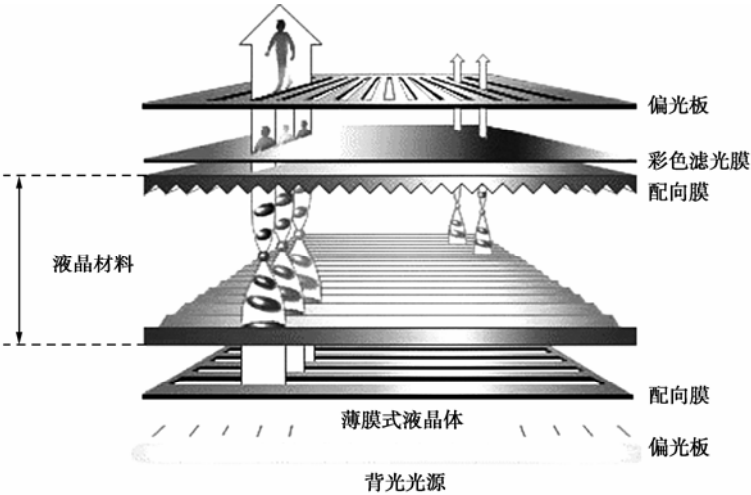


图 3.37 TFT 型液晶显示器原理示意图

与 CRT 显示器相比，液晶显示器具有工作电压低，没有辐射危害，功耗小，不闪烁，适于大规模集成电路驱动，体积轻薄（如图 3.38 所示），易于实现大画面显示和全色显示等特点，已经广泛应用于便携式计算机、数码摄像机、移动计算工具等设备，现在不少台式 PC 也配置了液晶显示器。



图 3.38 液晶显示器与 CRT 显示器外观比较

3. 主要性能参数

下面是 LCD 显示器和 CRT 显示器的一些主要性能参数。

(1) 显示屏的尺寸与比例

与电视机相同,计算机显示器屏幕的大小也是以显示屏的对角线长度来度量的,目前常用的显示器有 15 英寸、17 英寸、19 英寸等。显示屏的水平方向与垂直方向之比一般为 4:3。然而现在真正的电影一般都是宽银幕的,将宽银幕的电影转换为 4:3 总会造成画面质量、形状或者内容的损失。为了在计算机上更好地收看影视,宽屏的液晶显示器出现了,宽屏的特点就是屏幕的宽度明显超过高度。由于未来的高清晰电视主要会使用 16:9 的比例,因此目前越来越多的液晶显示器采用了宽屏的比例。不过 16:9 也有几个“变种”,比如 15:9 和 16:10,由于其比例和 16:9 比较接近,因此这 3 种屏幕比例的液晶显示器都可以称为宽屏。宽屏的比例更接近黄金分割比,也更适合人的眼睛,在观看影片时给人的感受也更舒服。

(2) 显示器的分辨率

分辨率是衡量显示器的一个重要指标,它指的是整屏可显示像素的多少,一般用水平分辨率 \times 垂直分辨率来表示,如 1024 \times 768, 1280 \times 1024 等。

(3) 刷新频率

刷新频率指所显示的图像每秒钟更新的次数。CRT 显示器将画面分成若干“扫描线”来进行刷新,CRT 内侧的荧光粉在接受电子束的轰击时,只能维持短暂的发光,根据人眼视觉暂留的特性,需要不断进行刷新(电子束重复轰击荧光粉)才能有稳定的视觉效果。CRT 的刷新频率要达到 85Hz 以上,才不闪烁。与 CRT 显示器相比,LCD 显示器产生图像不是通过电子枪扫描,而是采用灯管作为背光光源,通过控制液晶是否透光来控制亮和暗,每个像素都在持续发光,所以 LCD 的刷新是对整幅的画面进行刷新,LCD 即使在较低的刷新率(如 60Hz)下,也不会出现闪烁的现象,图像稳定。另外,从节省系统资源角度上考虑,在保证显示效果的前提下,将刷新率适当调低一些,也可以减轻显卡的负担。

(4) 带宽

带宽是代表显示器显示能力的一个综合指标,指每秒钟所扫描的像素个数,以 MHz 为单位。对 CRT 显示器而言,带宽越大表明显示控制能力越强,显示效果越佳。

带宽的详细计算公式如下:

理论上带宽=水平分辨率 \times 垂直分辨率 \times 刷新频率(即场频)

但在实际情况中,显像管电子束的扫描为了避免信号在扫描边缘的衰减,保证图像的清晰,扫描频率要比理论值高一些。所以,在计算带宽的时候还应该除以一个“有效扫描系数”,一般取值为 0.6~0.7。按照这个计算方法,分辨率为 1024 \times 768、刷新频率为 85Hz 的显示器,所需要带宽约为 106MHz。也就是说,显示器的带宽至少在 106MHz 以上。目前 CRT 显示器的最高带宽已达到 200MHz 以上,如美格 17 英寸 CRT 显示器中的 796FD II 和 796FDX5 的带宽都达到 203MHz。

由于刷新频率对于 LCD 显示器的重要性远低于 CRT,LCD 显示器的带宽,因此已经成为可有可无的技术指标。

(5) 可显示颜色数目

一个像素可显示出多少种颜色,由表示这个像素的二进位位数决定。彩色显示器的彩色是由 3 个基色 R, G, B 合成而得到的,因此是 R, G, B 3 个基色的二进位位数之和决定了可显示的颜色数目。例如,R, G, B 分别用 8 位表示,则它就有 $2^{24} \approx 1680$ 万种不同的颜色。

(6) 对比度

对比度定义为最大亮度值(全白)除以最小亮度值(全黑)的比值。随着对比度的提高,

显示器还原的色彩就越鲜艳，画面色彩的层次感更加分明，色阶过渡更细腻。人眼可以接受的对比度一般在 250:1 左右，日常使用的经验告诉我们，在绝大多数的情况下，对比度达到 350:1 就能够让人十分满意了。而 CRT 显示器可以轻易地达到 500:1 甚至更高。不过随着技术的进步，现在也推出了一些高对比度的液晶，如戴尔的 E228WFP 已经达到 800:1。目前普及型液晶的对比度基本上都在 300:1 以上。

(7) 亮度

亮度以 cd/m^2 （每平方米的烛光亮度）或者 nits 为单位，一般来说，亮度越高，人眼感觉色彩就越鲜艳、越舒适。如果显示器的亮度过低，各种颜色看起来感觉灰暗，不仅不能真实地还原图像的色彩，还会丢失许多图像细节。另外，在提高显示器亮度的同时，也要提高其对比度，否则就会出现整个显示屏发白的现象。LCD 的亮度不如 CRT，因为 LCD 属于被动性显示，像素本身无发光能力，是通过背光源来显像的，而 CRT 本身就具有发光显像性，并且 LCD 背光经过层层的光折损，能传至前方面板时已所剩无几，故透光率不高。

(8) 响应时间

响应时间是显示器各像素点对输入信号反应的速度，即像素由暗转亮或由亮转暗所需要的时间。响应时间越短则使用者在看动态画面时越不会有尾影拖曳的感觉。CRT 显示器响应时间仅为 1~3ms。所以，响应时间在 CRT 显示器中一般不会被人们提及，由于液晶显示器是利用液晶分子扭转来控制光的通断的，而液晶分子的扭转需要一个过程，所以 LCD 显示器的响应时间要明显长于 CRT。

LCD 显示器的响应时间的发展经过早期的 40ms, 25ms, 16ms, 8ms 到目前常见的 5ms 响应时间，应该说有了长足的进步。对于一般的用户来说，只要购买 8ms（1/0.008，每秒钟显示 125 帧画面）的产品已经可以基本满足日常应用的要求，对于游戏玩家而言，5ms 或更快的产品为较好的选择。

(9) 可视角度

可视角度是指用户可以从不同的方向清晰地观察屏幕上所有内容的角度，如图 3.39 所示。与 CRT 显示器接近 180° 的可视角度不同，由于提供液晶显示的光源经折射和反射后输出时已有一定的方向性，在超出这一范围观看就会产生色彩失真现象，因此 CRT 电视不会有这个问题。视角越大，观看的角度越好，LCD 显示器也就更具有适用性。目前市场上大多数产品的可视角度在 120° 以上，部分产品达到了 140° 以上。

(10) 接口

CRT 显示器因为设计制造上的原因，只能接收模拟信号输入，绝大多数使用的是 15 针 D-Sub 模拟信号接口。D-Sub 接口除了第 9 只针脚没有作用外（如图 3.40 所示），其余 14 只针脚各负责传递独立的模拟信号，包括 CRT 显示器所需要的 5 个基本分量（红、绿、蓝 3 色信号，垂直同步、水平同步信号）和用来同显示器通信的串行数据和串行时钟信号。

DVI (Digital Visual Interface) 数字显示接口，是近年来随着数字化显示设备的发展而发展起来的一种显示接口。

目前的 DVI 接口分为两种，一种是 DVI-D 接口，如图 3.41 所示，只能接收数字信号，接口上只有 3 排 8 列共 24 个针脚，其中右上角的一个针脚为空，不兼容模拟信号。另外一种则是 DVI-I 接口，如图 3.42 所示，可同时兼容模拟信号和数字信号。兼容模拟信号并不意味着模拟信号的接口 D-Sub 可以连接在 DVI-I 接口上，而必须通过一个转换接头才能使用。

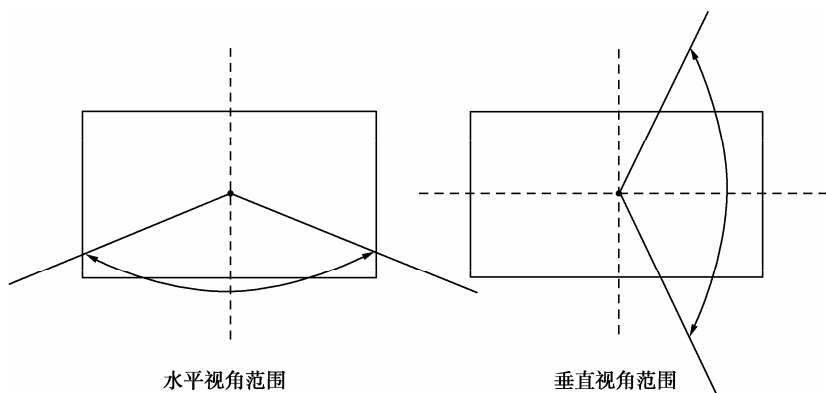


图 3.39 显示器可视角度示意图

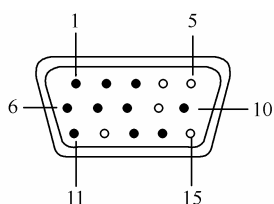


图 3.40 D-Sub 接口

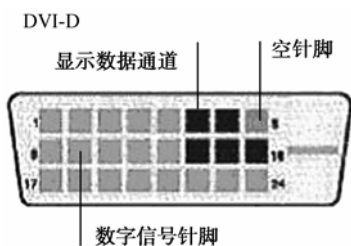


图 3.41 DVI-D 接口

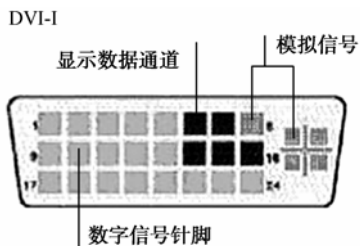


图 3.42 DVI-I 接口

普通的模拟显示接口在显示过程中，首先要在计算机的显卡中经过 D/A（数/模）转换，将数字信号转换为模拟信号传输到显示设备中，而在数字化显示设备中，又要经 A/D（模/数）转换将模拟信号转换成数字信号，然后显示。在经过 2 次转换后，不可避免地造成了一些信息的丢失，对图像质量也有一定影响。而在 DVI 接口中，计算机直接以数字信号的方式将显示信息传送到数字显示设备中，避免了 2 次转换过程，因此从理论上讲，采用 DVI 接口

的数字显示设备的图像质量要更好。另外，在 DVI 接口标准中还增加了一个热插拔监测信号，从而真正实现了即插即用，避免了在连接过程中需关闭计算机和显示设备的麻烦。现在大多数高档液晶显示器都采用了该接口，而一般的 LCD 显示器还是利用 D-SUB 来传递数据的。

（11）辐射和环保

随着科技不断发展，人们对健康和环保的要求越来越苛刻，对于电子产品，除了功能之外，同样也提出了这方面的要求。在显示器领域，TCO 认证就对显示器产品的环保提出了规范。2006 年 6 月，国际标准组织 TCO Development 发布了最新的显示器认证标准 TCO 06，如图 3.43 所示。TCO 认证是由瑞典专家委员会制定的世界上关于显示器环保要求的最严格的标准之一，要通过 TCO 认证，必须在生态（Ecology）、能源（Energy）、辐射（Emissions）及人体工学（Ergonomics）等 4 个方面都符合标准。



图 3.43 显示器认证标准 TCO 06 标志

【例 3.5】 飞利浦 107Q6 显示器（CRT）的性能参数如图 3.44 所示。



显示器尺寸：17 英寸
显示器接口：15 针 D-Sub
显像管类型：纯平
带宽：185MHz
栅距/水平：0.2mm
垂直扫描频率（这里指刷新频率）：50~160Hz
分辨率：1024×768
认证标准：TCO 03

图 3.44 飞利浦 107Q6 显示器（CRT）的性能参数

【例 3.6】 三星 225BW 显示器（LCD）的性能参数如图 3.45 所示。



液晶显示器屏幕尺寸：22 英寸
屏幕比例：16：10
分辨率：1680×1050
色彩：16.7M
液晶面板类型：TN
接口类型：D-Sub，DVI
平均亮度：280nits
点距：0.282mm
对比度：700：1
可视角度（水平/垂直）：160°/160°
响应时间：5ms

图 3.45 三星 225BW 显示器（LCD）的性能参数

4. 显示卡

计算机的显示系统包含显示器和显示卡（显卡）两部分，显卡的用途是将计算机系统所需要的显示信息进行转换，并向显示器提供行扫描信号，控制显示器的正确显示。PC 的显卡多半做成插卡的形式，它与显示器、CPU 及 RAM 的相互关系如图 3.46 所示。

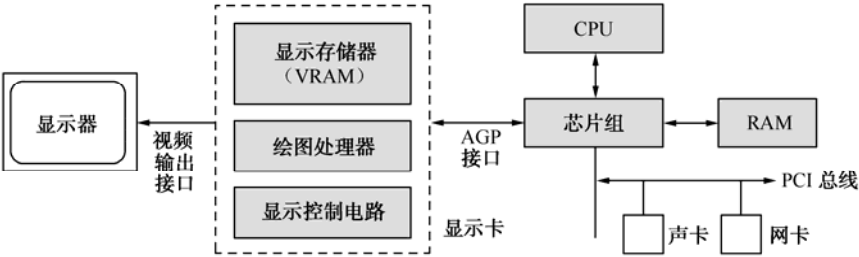


图 3.46 显卡、显示器、CPU 及 RAM 的关系

显卡主要由显示控制电路、图形处理器、显示存储器和接口电路 4 部分组成。

显卡的核心是图形处理器（Graphic Processing Unit，GPU）。早期的显卡比较简单，绘图功能都是由 CPU 在内存中完成的，然后将生成的图像数据由内存传送到显示存储器，显卡只起一个传输显示信息的作用。这样既降低了显示速度，又增加了 CPU 的工作量。随着 Windows 图形用户界面的出现，图形、图像应用的普及，这种弊端越来越严重，于是出现了图形处理器（又称为图形加速芯片）。这是一种专用处理器，它有一组可高速执行的适用于图像和图形处理的指令，如数据块传送、基本图形绘制、区域填色、图案填充、图形缩放、颜色转换等。由于采用硬件实现，所以速度很快，大大减轻了 CPU 的负荷，加快了图形操作速度。图形处理器可以按它们的数据并行传输带宽来进行划分，一般为 64 位或 128 位，更大的带宽可以使芯片在一个时钟周期内处理更多的信息。常见的生产显示芯片（图形处理器）的厂商包括 Intel, ATI, NVIDIA, VIA (S3), SIS, Matrox, 3D Labs。Intel, VIA (S3), SIS 主要生产集成芯片；ATI, NVIDIA 以独立芯片为主，是目前市场上的主流；Matrox, 3D Labs 则主要面向专业图形市场。

显示存储器又称为帧存储器、刷新存储器或简称为 VRAM（Video RAM），它用于存储显示屏上所有像素的颜色信息。在屏幕每次刷新时，都要将这些信息读出送到显示器去显示。VRAM 在物理上独立于 PC 的内存，但逻辑上是一个整体（与内存统一编址，CPU 可直接进行访问）。

显示控制电路负责对显卡的操作进行控制和协调，包括对 CRT（或液晶）显示器进行控制，如光栅扫描、同步、画面刷新等。

接口电路负责显卡与 CPU 和内存的数据传输。由于显卡在构造各种特殊效果时需要大量的内存来处理数据，而显卡上一般提供 64~128MB 的显示内存，大量的数据处理需要使用到系统内存，PCI 接口的显卡与内存之间的传输速率是 133Mbps，而显示内存与图形芯片之间的传输速率可以达到 1Gbps，这样就会影响显卡的显示速度。如果添加显示内存，则需要很大的费用，所以 Intel 为显卡和系统内存之间架起一条高速通道 AGP（Accelerated Graphics Port），是一个廉价的解决办法。AGP 显卡与 CPU、内存之间的数据传输速率远远超过了 PCI 显卡，其数据线宽度为 32 位，时钟频率为 66MHz，最高传输速率达 1Gbps（4 倍速）或 2.1Gbps（8 倍速），而且 AGP 显卡可以直接使用系统内存来处理图形。

【例 3.7】 丽台 A340 TD/128 显卡的性能参数如图 3.47 所示。



显卡核心: GeForce FX 5200
显卡属性: 桌面级
显卡核心频率: 250MHz
核心工艺: 150nm
核心生产厂商: NVIDIA
RAMDAC 频率: 350MHz
显存类型: DDR
显存频率: 400MHz
显卡显存容量: 128MB
显存位宽: 128 位

图 3.47 丽台 A340 TD/128 显卡的性能参数

随着 3D 图像技术和多媒体的广泛应用和深入，人们对显示质量提出越来越高的要求，各种更高性能的显示卡还将不断出现。

3.5.2 扫描仪

扫描仪（Scanner）是一种计算机输入设备，它可将各种图片、图纸、书稿等资料扫描输入到计算机中，转换成数字化图像数据进行保存和使用。配备专门的图像处理软件，计算机系统就可以进行图文档管理、图文排版、计算机广告创意、光学符号识别（OCR）、工程图纸扫描录入、计算机传真和复印等。

1. 扫描仪的分类

按扫描仪的结构来分，扫描仪可分为手持式、平板式、胶片和滚筒式等几种。

（1）手持式扫描仪

手持式扫描仪在工作时，需要操作人员用手拿着扫描仪在原稿上移动。它的扫描头比较窄，只适用于扫描小幅图稿，如照片。不同型号的手持式扫描仪的扫描速度目前大致相当，实际的扫描速度与所使用的主机速度和软件有关，手持式扫描仪用手拖动时不能过快，否则就要丢失图像数据。



图 3.48 使用广泛的平板式扫描仪

（2）平板式扫描仪

平板式扫描仪是应用最广泛的扫描仪，如图 3.48 所示。平板式扫描仪在扫描时，图稿平铺在台面上，由步进电机带动扫描头作直线运动进行扫描。平板式扫描仪的工作性能比较可靠、使用寿命长，产品安装、使用方便，价格也较便宜，已经在家用和办公自动化领域得到广泛应用。

（3）胶片扫描仪和滚筒式扫描仪

胶片扫描仪和滚筒式扫描仪都是高分辨率的专业扫描仪，它们在光源、色彩捕捉等方面均具有较高的技术性能，光学分辨率很高。这种扫描仪多数应用于专业印刷排版领域。

2. 扫描仪的工作原理

扫描仪是基于光电转换原理而设计的，上述几种类型的扫描仪工作原理大体相同，只

不过是结构和使用的感光器件不同而已。下面以平板式扫描仪为例，介绍它的工作原理，如图 3.49 所示。

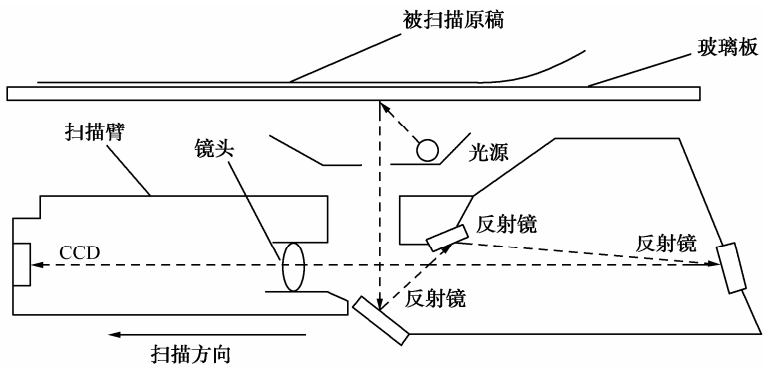


图 3.49 平板式扫描仪的工作原理

扫描仪工作时，将被扫描的原稿正面朝下放置在扫描仪玻璃板上。扫描仪采用高密度的光束照射图像，由电机牵动的扫描头沿着原稿移动，并接收从原稿上反射回来的光束。由于色彩及灰度的区别，反射回来的光强度也不同，这种反射光被聚焦后照射在 CCD（Charge Coupled Device，电荷耦合器件）上，通过光电转换产生电流输出。照射光强，电流大，照射光弱，电流小，再经模/数转换器（A/D 转换器）转换，就变成计算机可以处理的数字信号。这种数字信号还要由专门的软件进行各种校正和平滑处理，得到的图像数据以指定的文件格式（如 TIF 文件）存储在计算机中。

3. 扫描仪的性能指标

扫描仪的性能指标主要有分辨率、色彩位数、扫描速度、扫描幅面等。

（1）分辨率

分辨率反映了图像扫描仪的扫描精度，通常用每英寸上对图像的采样点来表示，标记为 dpi（dot-per-inch）或 ppi（pixel-per-inch）。从物理上讲，分辨率就是扫描仪 CCD 的排列密度，如 1000dpi 就表示该扫描仪的 CCD 排列密度为每英寸有 1000 个 CCD 器件。现在扫描仪还采用插值算法来进一步提高其分辨率，为了便于区别，人们把 CCD 的密度称为光学分辨率或物理分辨率。

（2）色彩位数

色彩位数（色彩深度）反映了扫描仪对图像色彩的辨析能力，色彩位数越多，扫描仪所能反映的色彩就越丰富，扫描的图像效果也越真实。色彩位数可以是 24 位、30 位、42 位、48 位等。

（3）预扫时间

预扫时间其实就是扫描仪对所有的扫描面积进行一次快速扫描所需的时间。扫描仪在开始扫描稿件时，必须通过预扫的步骤确定稿件在扫描平台上的位置。

（4）扫描速度

扫描速度的表示方式一般有两种：一种用扫描标准 A4 幅面所用的时间来表示，另一种使用扫描仪完成一行扫描的时间来表示。一般情况下，扫描黑白、灰度图像，扫描速度为 2~100ms/线，扫描彩色图像，扫描速度为 5~200ms/线。

(5) 扫描幅面

扫描幅面是指扫描仪允许被扫描原稿的最大尺寸，如 A4、A4 加长、A3、A1、A0 等。

除以上指标外，用户还要考虑与主机的接口形式、操作环境、随机软件是否丰富，以及安装的方便性等因素。

【例 3.8】 佳能 CanoScan 4400F 扫描仪的性能参数如图 3.50 所示。



扫描仪类型：平板式
扫描元件：6 线彩色 CCD
最大幅面：A4
光学分辨率：4800×9600 位
色彩位数：48 位
扫描速度：14P/min（页/分钟）
预扫时间：5s
扫描介质：图片,文字,照片,35mm 条带(负片/正片)/
幻灯片
接口：USB 2.0

图 3.50 佳能 CanoScan 4400F 扫描仪的性能参数

4. 图像扫描工具软件

软件在今天扫描仪的技术中所占的比重越来越大，尽管几乎所有的扫描仪都提供了扫描仪应用程序，但是用户可以使用很多种其他的标准图像处理软件来控制扫描仪扫描图片。这样做的意义在于，用户可以使用自己熟悉的图像工具来操作，而不必另外安装多余的软件。有些扫描软件中直接集成了 OCR（即文字识别软件）功能，紫光 OCR 系列和尚书系列都是捆绑扫描仪一起销售的软件，配合双分辨率功能，使扫描仪的易用性大大提高。用户不必在遇到文字时单独启动 OCR 软件进行文字部分的扫描，扫描仪会自动对文字部分采用合适的分辨率进行扫描，比如对文字进行 300dpi 扫描，而同时对图像部分进行 1200dpi 扫描，从而极大地提高了扫描速度和录入效率。此外，一些产品将多字体识别和字体颜色识别技术与 OCR 技术结合在一起工作，使扫描文档在计算机中保持硬复制文档的原貌。

3.5.3 数码相机

数码相机（Digital Camera）（如图 3.51 所示）是扫描仪之外的另一种重要的图像输入设备。数码相机又叫数字相机，与传统照相机相比，它不需要胶卷和暗房，能直接将照片以数字形式记录下来，并输入计算机中进行存储、处理和显示，或通过打印机打印出来。



图 3.51 数码相机的外观

1. 数码相机的工作原理

数码相机是由镜头、CCD、A/D（模/数转换器）、MPU（微处理器）、内置存储器、LCD（液晶显示器）、PC 卡（可移动存储器）和接口（计算机接口、电视机接口）等部分组成的，数码相机中只有镜头的作用与普通相机相同，其他部分则完全不同。

数码相机在工作时外部景物通过镜头将光线会聚到成像芯片（CCD 或 CMOS）上，并由成像芯片转换成电信号，再经模/数转换（A/D 转换）变成数字图像，接下来 MPU（微处理器）对数字信号进行必要的图像处理和数据压缩（大多采用 JPEG 标准）后，存储在相机内的内置存储器中。至此，数码相机的主要工作已经完成，使用者可通过 LCD 查看拍摄的照片。数码相机为扩大存储容量可使用可移动存储器，如 PC 卡或 CF 卡。此外，数码相机还提供了连接到计算机和电视机的接口，如图 3.52 所示。

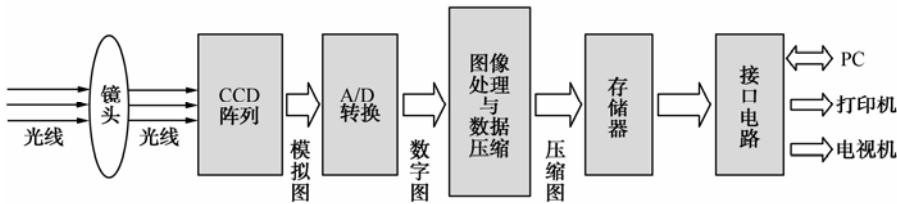


图 3.52 数码相机工作原理图

2. 数码相机的性能指标

(1) 像素数目

在使用数码相机拍摄的过程中，由于接收信号的感光元件是 CCD 芯片，所以厂商经常用 CCD 芯片的像素作为数码相机的规格。

CCD 芯片中有大量的 CCD 像素，每一个像素可记录图像中的一个点，然后将其光信号转换为电信号。显然，CCD 像素越多，影像分解的点就越多，最终所得到的影像的分辨率（清晰度）就越高，图像的质量也就越好。所以，CCD 像素的数目是数码相机的一个至关重要的性能指标。

选用多少像素的数码相机合适，完全取决于对最终图像大小和清晰度的要求。例如，获取供计算机显示或网页上使用的图像，以及制作 3~5 英寸以下的照片时，中低分辨率（1024×768~1600×1200）即可满足要求。所以 200~300 万像素左右的数码相机，已经能满足普通消费者的一般使用要求。

(2) 存储容量

经过 CCD 芯片成像并转换得到的数字图像，存储在数码相机的内置存储器中。数码相机的存储器大多采用快擦除存储器（闪存），即使断电也不会丢失信息。存储容量是数码相机的另一项重要性能指标，在图像分辨率和质量要求相同的情况下，存储容量越大，可存储的数字相片就越多。

(3) 接口

数码相机上常见的接口主要有 3 种，它们几乎在所有数码相机上都可以看到。

① 外接电源接口：主要用于数码相机在室内长时间工作时的供电，部分机型还可以通过这一接口对电池进行充电；不过有少部分机型，这一接口就只能用来充电。

② 外接视频输出接口：通过这一接口可以将数码相机上的图像在电视机或其他视频设备上播放。

③ USB 接口：通用串行总线接口。利用这一接口，我们不但可以把照片从相机中复制到计算机上，还可以将相机当摄像头使用，部分数码相机还可以通过这一接口直接连接上打印机打印出照片。大部分机型支持 USB 1.1 接口（1.5MB/s），现在支持 USB 2.0 接口（60MB/s）的机型也渐渐多了起来。部分数码相机还支持 IEEE 1394 火线接口、IRDA（Infrared Data Association）红外接口。

目前数码相机的结构已日趋完善，功能趋于多样化。一般使用的轻便数码相机结构上都配置有彩色液晶显示器、USB 数字接口和模拟视频信号输出接口，具有自动聚集、自动曝光、自动白平衡调整、数字变焦、影像预视、影像删除等功能，有的还增设了连续拍摄功能，可满足人们多样化的需求。

数码相机使用的成像芯片目前采用 CCD 芯片居多数，像素在 200~300 万以下的普及型相机大多采用 CMOS 成像芯片，价格比较便宜。

【例 3.9】 索尼 T10 数码相机的性能参数如图 3.53 所示。



有效像素数：720 万
最大像素数：740 万
传感器尺寸：1/2.5 英寸
传感器类型：CCD
存储卡类型：MS Duo 卡
内置/随机存储卡容量：58MB
图像尺寸：3072×2304, 3072×2048, 2592×1944, 2048×1536,
1632×1224, 1920×1080, 640×480
音频输入：有
照片格式：JPEG
视频格式：MPEG
连接类型：USB 2.0

图 3.53 索尼 T10 数码相机的性能参数

3.5.4 打印机

打印机也是 PC 的一种主要输出设备，它把程序、数据、字符、图形打印在纸上。目前使用较广的打印机有激光打印机和喷墨打印机两种。

1. 打印原理

(1) 激光打印机

激光打印机是激光技术与复印技术相结合的产物，它是一种高质量、高速度、低噪声、价格适中的输出设备，如图 3.54 所示。

激光打印机由激光器、旋转反射镜、聚焦透镜和感光鼓等几部分组成。激光器采用半导体激光二极管，在它的两极加上大小不同的电压就会发出强度变化的激光束，这称为电源调制。计算机输出的“0”，“1”信号加在激光二极管上，就能得到一系列被调制的脉冲式激光。激光束经棱镜反射后聚焦到感光鼓，感光鼓表面涂有光电转换材料，于是，计算机输出的文字或图形就以不同密度的电荷分布记录在感光鼓表面，以静电形式形成了“潜像”。



(a) 黑白激光打印机



(b) 彩色激光打印机

图 3.54 激光打印机

墨粉是一种带电荷的细微塑料颗粒，其电荷与感光鼓表面的电荷极性相反。当带有电荷的感光鼓表面经过涂墨辊时，有电荷的部位就吸附了墨粉颗粒，“潜像”就变成了真正的影像。在感光鼓转动的同时，另一组传动系统将打印纸送进来，经过一组电极，打印纸带上了与感光鼓表面极性相同但强得多的电荷，随后纸张经过带有墨粉的感光鼓，感光鼓表面的墨粉被吸附到打印纸上，图像就在纸张表面形成了。此时，墨粉和打印纸仅仅是靠电荷的引力结合在一起的，打印纸被送出打印机之前，经过高温加热，塑料质的墨粉被熔化，在冷却后固着在纸张表面。将墨粉传给打印纸之后，感光鼓表面继续旋转，经过一个清洁器，将剩余的墨粉去掉，以便进入下一个打印循环。由于激光能聚焦成很细的光点，因此激光打印机的分辨率较高，印刷质量相当好。激光打印机的工作原理如图 3.55 所示。

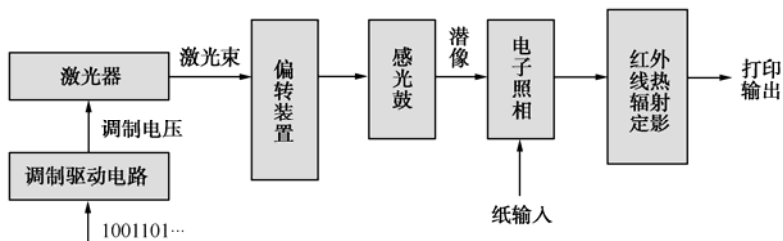


图 3.55 激光打印机的工作原理

激光打印机大部分使用并行接口或 USB 接口，一些高速激光打印机使用 SCSI 接口。

激光打印机分为黑白和彩色两种，黑白激光打印机使用黑色墨粉来印刷，彩色激光打印机采用青、品红、黄、黑 4 种墨粉 (CMYK)，依靠颜色混合就形成了丰富的色彩。低速黑白激光打印机已经普及，而彩色激光打印机的价格还比较高，适合专业用户使用。

(2) 喷墨打印机

喷墨打印机的优点是能输出彩色图像，经济、低噪声，打印效果好，使用低电压不产生臭氧，有利于保护办公室环境等。在彩色图像输出设备中，喷墨打印机已占绝对优势，如图 3.56 所示。

喷墨打印机按打印头的工作方式可以分为压电喷墨技术和热喷墨技术两大类。按照喷墨材料的性质又可以分为水质料、固态油墨和液态油墨等类型。



图 3.56 彩色喷墨打印机与墨盒

压电喷墨技术是将许多小的压电陶瓷放置到喷墨打印机的打印头喷嘴附近，利用它在电压作用下会发生形变的原理，适时地给它加压，压电陶瓷随之产生伸缩，使喷嘴中的墨汁喷出，在输出介质（打印纸）表面形成图案，如图 3.57（a）所示。采用压电喷墨技术可以通过控制电压来调节墨滴的大小，从而获得较高的打印精度和打印效果。

热喷墨技术（如 Canon 和 HP 等公司的产品）是让墨水通过细喷嘴，在强电场的作用下，将喷头管道中的一部分墨汁气化，形成一个气泡，气泡将喷嘴处的墨水顶出，以每秒数千次的频率喷射到输出介质表面，形成图案或字符，如图 3.57（b）所示。这种喷墨打印机有时又被称为气泡打印机，用这种技术制作的喷头工艺比较成熟，成本也较低廉。

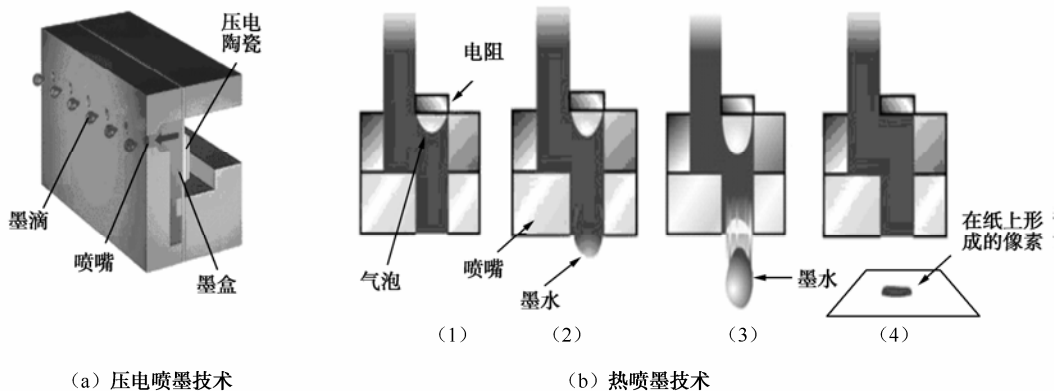


图 3.57 两种喷墨技术

喷墨打印机的关键技术是喷头。要使墨水从喷嘴中以每秒近万次的频率喷射到纸上，这对喷嘴的制造材料和工艺要求很高。喷墨打印机所使用的耗材是墨水，理想的墨水应不损伤喷头，能快干但又不在于喷嘴处结块，防水性好，不在纸张表面扩散或产生毛细渗透现象，在普通纸张上打印效果要好，不因纸张种类不同而产生色彩偏移现象，黑色要纯，色彩要艳，图像不会因日晒或久置而褪色，墨水应无毒，不污染环境，不影响纸张再生使用。由于有上述许多要求，因此墨水成本高，而且消耗快，这是喷墨打印机的不足之处。

2. 打印机的性能指标

打印机的性能指标主要是打印精度、打印速度、色彩数目和打印成本等。

(1) 打印精度

打印精度也就是打印机的分辨率，它用 dpi（每英寸可打印的点数）来表示，是衡量图像清晰程度最重要的指标。300dpi 是人眼分辨文本与图形边缘是否有锯齿的临界点，再考虑到其他一些因素，因此 360dpi 以上的打印效果才能基本令人满意。激光打印机的分辨率最低为 300dpi，有的产品为 400dpi，600dpi，800dpi，甚至达到 1200dpi。喷墨打印机的分辨率一般可达 300~360dpi，高的能达到 1000dpi 以上。

(2) 打印速度

激光打印机和喷墨打印机是一种页式打印机，它们的速度单位是每分钟打印多少页纸（P/min），家庭用的低速打印机大约为 4P/min，办公使用的高速打印机速度可达到 10P/min 以上。

(3) 色彩数目

这是指打印机可打印的不同彩色的总数。对于喷墨打印机来说，最初只使用 3 色墨盒，色彩效果不佳。后来改用青、品红、黄、黑 4 种墨水（CMYK），虽然有了很大改善，但与专业要求相比还是不太理想。为了提高彩色输出质量，先后出现过许多先进技术。作为 Canon 公司的“Photo Realism”技术的一部分，在原来 4 色墨水的基础上，再创 3 种崭新的 Photo 墨水（Photo 黄、Photo 品红、Photo 青），从而实现了 7 重色打印，进一步丰富了打印图片或图像的颜色层次。Canon 公司独创了 25 重色调打印技术，该技术利用 7 种颜色墨水，可打印组合出多达 25 重层次的色调，可以更加清晰地表现颜色中的细微变化，完美再现图像的中间色调，从而打印出品质更高的图像。

(4) 其他性能指标

包括打印成本、噪声、打印幅面大小、可打印字体的数目及种类、功耗及节能功能、可打印的副本数目、与主机的接口类型等。

【例 3.10】 HP LaserJet 1320 黑白激光打印机的性能参数如图 3.58 所示。



打印机类型：黑白激光打印机
适用类型：家用打印机/商用打印机
硒鼓型号：Q5949A
硒鼓寿命：2500 张
缓存：16MB
接口：USB 2.0/并口
最高分辨率：1200×1200dpi
黑白打印速度：21P/min
最大打印幅面：A4
供纸方式：自动/手动
纸张容量：250 张
首页出纸时间：8s
噪声：62dB

图 3.58 HP LaserJet 1320 黑白激光打印机的性能参数

【例 3.11】 HP Deskjet D1368 彩色喷墨打印机的性能参数如图 3.59 所示。



打印机类型：彩色喷墨打印机
适用类型：家用打印机
墨盒支持：C8816A, C8817A
墨盒类型：四色墨盒
接口：USB 2.0
字体：内置字体
最高分辨率：4800×1200dpi
黑白打印速度：20P/min
彩色打印速度：16P/min
大幅面打印速度：12P/min
最大打印幅面：A4
供纸方式：自动/手动
纸张容量：80 张
噪声 54dB

图 3.59 HP Deskjet D1368 彩色喷墨打印机的性能参数

3.6 数字图像处理

数字图像处理（Digital Image Processing）又称为计算机图像处理，它是指将图像信号转换成数字信号并利用计算机对其进行处理的过程。数字图像处理最早出现于 20 世纪 50 年代，当时的电子计算机已经发展到一定水平，人们开始利用计算机来处理图形和图像信息。数字图像处理作为一门学科大约形成于 20 世纪 60 年代初期。早期的图像处理的目的是改善图像的质量，它以人为对象，以改善人的视觉效果为目的。在图像处理中，输入的是质量低的图像，输出的是改善质量后的图像，常用的图像处理方法有图像增强、复原、编码、压缩等。随着图像处理技术的深入发展，从 20 世纪 70 年代中期开始，由于计算机技术和人工智能、思维科学研究的迅速发展，使数字图像处理向更高、更深层次发展。

3.6.1 数字图像处理的内容

数字图像处理主要研究的内容有以下几个方面。

1. 图像变换

由于图像阵列很大，直接在空间域中进行处理，涉及计算量很大。因此，往往采用各种图像变换的方法，如傅里叶变换、沃尔什变换、离散余弦变换（DCT）等间接处理技术，将空间域的处理转换为变换域处理，不仅可减少计算量，而且可获得更有效的处理（如傅里叶变换可在频域中进行数字滤波处理）。目前新兴研究的小波变换在时域和频域中都具有良好的局部化特性，它在图像处理中也有着广泛而有效的应用。图像变换还包括传统的几何变换，如图像的缩放、旋转、平移、投影变换等。

2. 图像编码压缩

图像编码压缩技术可减少描述图像的数据量（即比特数），以便节省图像传输、处理时间并减少所占用的存储器容量。压缩可以在不失真的前提下获得，也可以在允许的失真条件下进行。编码是压缩技术中最重要的方法，它在图像处理技术中是发展最早且比较成熟的技

术（详见第 5 章）。

3. 图像增强和复原

图像增强和复原的目的是为了提高图像的质量，如去除噪声，提高图像的清晰度等。图像增强不考虑图像质量下降的原因，只突出图像中所感兴趣的部分，如图 3.60 所示。如强化图像高频分量，可使图像中物体轮廓清晰，细节明显；如强化低频分量可减少图像中的噪声影响。

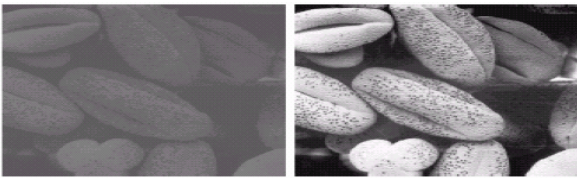


图 3.60 图像增强

图像复原要求对图像质量下降的原因有一定的了解，一般讲应根据质量下降过程建立“降质模型”，再采用某种滤波方法，恢复或重建原来的图像，如图 3.61 所示。



图 3.61 图像复原

4. 图像分割

图像分割是数字图像处理中的关键技术之一。图像分割是将图像中有意义的特征部分提取出来，如图像中的边缘、区域等，这是进一步进行图像识别、分析和理解的基础，如图 3.62 所示。虽然目前已研究出不少边缘提取、区域分割的方法，但还没有一种普遍适用于各种图像的有效方法。因此，对图像分割的研究还在不断深入之中，是目前图像处理中研究的热点之一。



图 3.62 图像分割

5. 图像描述

图像描述是图像识别和理解的必要前提。经过图像分割后，得到的是若干区域和边界，通常把感兴趣的部分叫目标，其余的部分叫背景，最终目的是要让计算机来识别这些目标。为了有效地识别目标，必须描述这些目标即提供它们的有用信息和相互关系。一般图像的描述方法采用二维形状描述，它有边界描述和区域描述两类方法。对于特殊的纹理图像可采用二维纹理特征描述。随着图像处理研究的深入发展，已经开始进行三维物体描述的研究，提出了体积描述、表面描述、广义圆柱体描述等方法。

6. 图像分类（识别）

图像分类（识别）属于模式识别的范畴，其主要内容是图像经过某些预处理（增强、复原、压缩）后，进行图像分割和特征提取，从而进行判决、分类。图像分类常采用经典的模式识别方法，有统计模式分类和句法（结构）模式分类，近年来新发展起来的模糊模式识别和人工神经网络模式分类在图像识别中也越来越受到重视。常见的识别应用如下。

1) 光学符号识别（OCR）

OCR（Optical Character Recognition）技术实际上是计算机认字，它通过扫描和摄像等光学输入方式获取纸张上的文字图像信息，利用各种模式识别算法分析文字形态特征，判断出文字的标准码，并按通用格式存储在文本文件中。所以 OCR 是一种非常快捷、省力的文字输入方式，也是在文字数据量大的今天被人们广泛采用的输入方法。汉字识别 OCR 就是使用扫描仪对输入计算机的文本图像做识别，自动产生汉字文本文件，采用 OCR 与人工输入的汉字效果是一样的，但速度比手工快几十倍甚至上百倍，如图 3.63 所示。所以现在的扫描仪都附带 OCR 识别软件，如尚书、紫光、汉王等。

一个 OCR 系统可分为如下 3 个部分。

(1) 预处理部分

首先把待识别的文本通过扫描设备输入系统，由硬件、软件完成数字图像处理，把待识别文本中的照片、图形与文字分离开来，并将分离出来的文字分割成单个符号图形供识别部分使用。

(2) 识别部分

把分割出的文字图形规格化，提取文字的几何特征和统计特性，并把特征送入识别器，将得到的待识文字的内码作为结果。

(3) 后处理部分

对识别结果及预处理部分的某些因素进行综合考虑，生成具有一定格式的识别结果，对整个识别结果进行语言学方面的检查，纠正误识部分，从而产生 OCR 系统对该识别文本的最终结果。

2) 手写识别

手写识别是指将在手写设备上书写时产生的有序轨迹信息转化为汉字内码的过程，实际

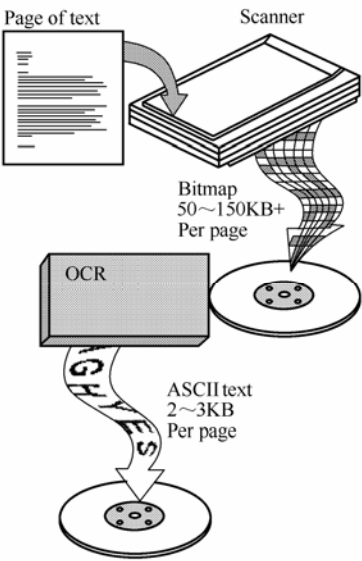


图 3.63 通过 OCR 输入文本

上是手写轨迹的坐标序列到汉字内码的一个映射过程。

目前的手写识别技术一般要经过数据采集、预处理、归一化、特征抽取、特征匹配及输出文字代码几个主要阶段。其中，数据采集一般通过手写板或触摸屏等输入设备实现，如图 3.64 所示。用户在这些设备上书写的笔画以类似于矢量图的形式存储下来。由于采集到的信息一般含有较高的噪声，因此在提取它的特征信息前先要进行预处理，将干扰信息尽量剔除。



图 3.64 手写输入设备

做完了这些前期的铺垫工作，就可以进入关键的特征抽取阶段了。手写留下的轨迹图中一些关键的信息将被提取出来作为识别的依据，这些特征信息包括笔画的长短、角度及各笔画的交叉点、组成结构等。这些抽取出来的特征信息与系统内建的一个识别字典相对照，便得到了识别的结果。

即便是同一个人写同一个汉字，其形态也会有所差异，因此最后抽取出来的信息通常不能与识别字典里的特征信息完全吻合。根据吻合程度不同，识别结果会有多个备选字。手写输入系统通常将它们按照匹配程度的顺序列出，供用户选择。

(1) 手写识别软件

手写识别软件（如图 3.65 所示）是手写输入的核心部分，它决定了汉字输入的识别率及汉字输入的易用性和可操作性。

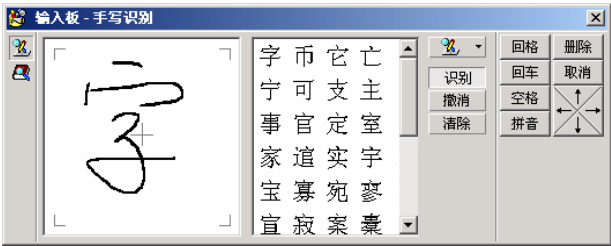


图 3.65 微软拼音输入法 2003 自带的手写识别软件

大体上可从以下几个方面来考察识别软件。

- ① 连笔识别。每人写字都希望既快又准确，而汉字笔画的连笔较多，一些软件能有效地识别一些行书和草书，以便提高输入的效率。
- ② 自学习功能。使手写识别更能适应非特定人的非特定书写，使用者可将自己的一些独特运笔而生成的文字，通过此功能让计算机识别记忆下来，下次再遇到相同的字，计算机就能轻松识别了。
- ③ 联想字识别、同音字识别、同形字识别等技术。这些功能能够有效解决如提笔忘字、写错别字等问题，以便提高效率，增加输入速度。

(2) 手写识别与 OCR 识别的异同

从技术上讲，手写识别与 OCR 具有一定的相似之处，但两者仍存在较大的区别。

首先，OCR 的信息源是一个点阵图形，各个点之间没有明确的联系，也没有时间的先后。而手写输入采集到的是矢量图，构成图形的各个点之间有着较明确的联系，也有时间的先后顺序。由于 OCR 面对的是一幅点与点之间没有关联性的点阵图，因此，它比手写识别的技术难度更高。

其次，手写字通常是一种不规范的字体，而 OCR 面对的大多是较规范的印刷字体。手写识别目前的一个研究重点是连笔识别，它通常比逐笔写的汉字字体形变更大，而且连笔手写每个人的书写差异也更大，会产生很多的变化。

3) 指纹识别和人脸识别

指纹识别和人脸识别都属于生物特征识别技术，与传统的身份鉴定手段相比，生物特征识别技术具有以下特点：①生物特征是人体所固有的特征，随身携带，不易遗忘或丢失，使用非常方便；②生物特征与人体是唯一绑定的，防伪性好，不易伪造或被盗，没有任何两个手指指纹的纹线形态一致，指纹纹线的形态终生不变，如图 3.66 (a) 所示，人脸具有唯一性，如图 3.66 (b) 所示。

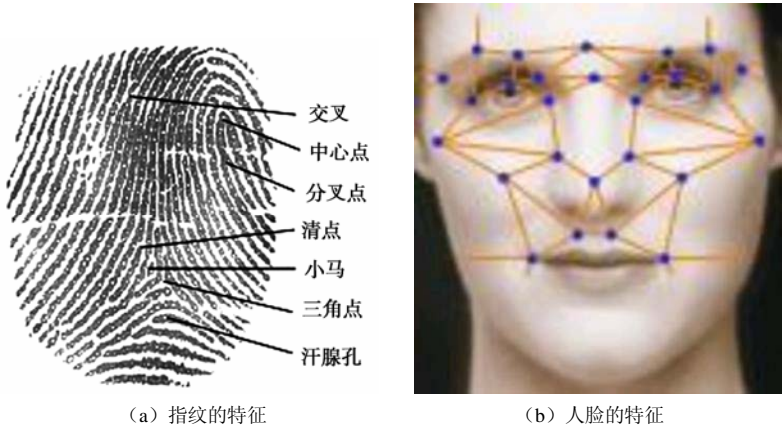


图 3.66 指纹和人脸的特征

指纹识别由两个过程组成，即登记过程和识别过程，原理框图如图 3.67 所示。

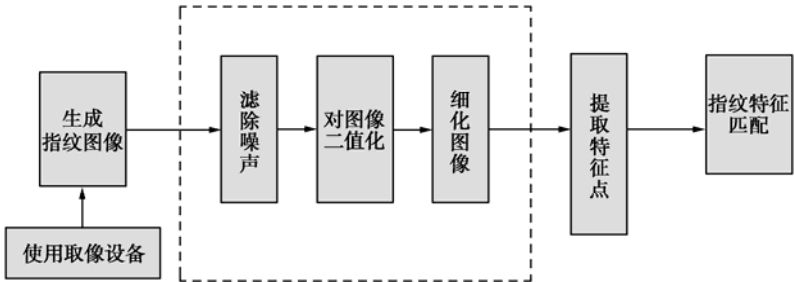


图 3.67 指纹识别原理框图

在登记过程中，用户需要先采集指纹，然后计算机系统将自动进行特征提取，提取后的特征将作为模板保存在数据库或其他指定的地方。

在识别或验证阶段，用户首先也要采集指纹，然后计算机系统将自动进行特征提取，提取后的特征将与数据库中的模板进行比对，给出比对结果。

此过程是一个通用的过程，对所有的生物特征识别技术都适用（包括人脸识别）。

近几年来，全球的生物识别技术已从研究阶段转向应用阶段，对该技术的研究和应用如火如荼，前景十分广阔，生物识别技术正越来越多地影响着人们的日常生活。通过取代个人识别码和口令，生物识别技术不仅可以阻止非授权访问，还能防止盗用 ATM、电话、智能卡、桌面 PC、笔记本电脑、工作站及计算机网络。生物识别技术也可在电话、网络进行金融交易时进行身份认证，或在办公场所取代现有的钥匙、证件、印章等。

3.6.2 数字图像处理的应用

图像是人类获取和交换信息的主要来源，因此，图像处理的应用领域必然涉及人类生活和工作的方方面面。随着人类活动范围的不断扩大，图像处理的应用领域也将随之不断扩大。

1. 航天和航空技术方面的应用

数字图像处理技术在航天和航空技术方面得到了应用。首次获得实际成功应用的是美国喷气推进实验室（JPL），他们对航天探测器徘徊者 7 号在 1964 年发回的几千张月球照片使用了图像处理技术，如几何校正、灰度变换、去除噪声等，并考虑了太阳位置和月球环境的影响，由计算机成功地绘制出月球表面地图，获得了巨大的成功。随后又对探测飞船发回的近十万张照片进行更为复杂的图像处理，以至获得了月球的地形图、彩色图及全景镶嵌图，获得了非凡的成果，为人类登月创举奠定了坚实的基础，也推动了数字图像处理这门学科的诞生。在以后的宇航空间技术，如对火星、土星等星球的探测研究中，数字图像处理技术都发挥了巨大的作用。除了对月球、火星照片的处理之外，另一方面的应用是在飞机遥感和卫星遥感技术中。许多国家每天派出很多侦察飞机对地球上感兴趣的地区进行大量的空中摄影。对由此得来的照片进行处理分析，以前需要雇佣几千人，而现在改用配备有高级计算机的图像处理系统来判读分析，既节省人力，又加快了速度，还可以从照片中提取人工所不能发现的大量有用情报。从 20 世纪 60 年代末以来，美国及一些国际组织发射了资源遥感卫星（如 LANDSAT 系列）和天空实验室（如 SKYLAB），由于成像条件受飞行器位置、姿态、环境条件等影响，图像质量总不是很高。因此，用如此昂贵的代价进行简单直观的判读来获取图像是不合算的，必须采用数字图像处理技术。现在世界各国都在利用陆地卫星所获取的图像进行资源调查（如森林调查、海洋泥沙和渔业调查、水资源调查等），灾害检测（如病虫害检测、水火检测、环境污染检测等）、资源勘察（如石油勘察、矿产量探测、大型工程地理位置勘探分析等）、农业规划（如土壤营养、水分和农作物生长、产量的估算等）、城市规划（如地质结构、水源及环境分析等）。

2. 生物医学工程方面的应用

数字图像处理在生物医学工程方面的应用十分广泛，而且很有成效。1972 年，英国 EMI 公司工程师 Housfield 发明了用于头颅诊断的 X 射线计算机断层摄影装置，也就是我们通常所说的 CT（Computer Tomograph）。CT 的基本方法是根据人的头部截面的投影，经计算机处理来重建截面图像，称为图像重建。1975 年，EMI 公司又成功研制出全身用的 CT 装置，获得了人体各个部位鲜明清晰的断层图像。1979 年，这项无损伤诊断技术获得了诺贝尔奖，说

明它对人类做出了划时代的贡献。除了 CT 技术之外，还有一类是对医用显微图像的处理分析，如红细胞、白细胞分类，染色体分析，癌细胞识别等。此外，在 X 光肺部图像增晰、超声波图像处理、心电图分析、立体定向放射治疗等医学诊断方面都广泛地应用图像处理技术。

3. 通信工程方面的应用

当前通信的主要发展方向是声音、文字、图像和数据结合的多媒体通信。具体地讲，是将电话、电视和计算机以三网合一的方式在数字通信网上传输。其中，以图像通信最为复杂和困难，因图像的数据量十分巨大，如传送彩色电视信号的速率达 100Mbps 以上。要将这样高速率的数据实时传送出去，必须采用编码技术来压缩信息的比特量。从一定意义上讲，编码压缩是这些技术成败的关键。除了已应用较广泛的熵编码、DPCM 编码、变换编码外，目前国内外正在大力开发研究新的编码方法，如分形编码、自适应网络编码、小波变换图像压缩编码等。

4. 工业和工程方面的应用

在工业和工程领域中图像处理技术有着广泛的应用，如自动装配线中检测零件的质量，并对零件进行分类；印制电路板瑕疵检查；弹性力学照片的应力分析；流体力学图片的阻力和升力分析；邮政信件的自动分拣；在一些有毒、放射性环境内，识别工件及物体的形状和排列状态；在先进的设计和制造技术中采用机器视觉等。其中值得一提的是研制具备视觉、听觉和触觉功能的智能机器人，将会给工农业生产带来新的激励，目前已在工业生产中的喷漆、焊接、装配中得到有效的利用。

5. 军事公安方面的应用

在军事方面图像处理和识别主要用于导弹的精确末制导，各种侦察照片的判读，具有图像传输、存储和显示的军事自动化指挥系统，飞机、坦克和军舰模拟训练系统等；公安业务图片的判读分析，指纹识别，人脸鉴别，不完整图片的复原，以及交通监控、事故分析等。目前已投入运行的高速公路不停车自动收费系统中的车辆和车牌的自动识别都是图像处理技术成功应用的例子。

6. 文化艺术方面的应用

目前这类应用有视频图像的编辑，动画的制作，电子游戏的制作，纺织工艺品设计，服装设计与制作，发型设计，文物资料照片的复制和修复，运动员动作分析和评分等，现在已逐渐形成一门新的艺术——计算机美术。

3.7 图像处理工具软件

图像处理的工具软件很多，目前常用的有 Corel Systems 公司的 Photo Paint、Aldus 公司的 Photo Styler、Adobe 公司的 Photoshop 等。在 Windows 环境下，这些图像应用软件在其功能上都具有一定的共性，包括：

- (1) 支持多种图像数据格式，具有图像编辑、变形变换、优化处理等功能。
- (2) 可选定某个区域进行裁剪、复制、粘贴、水平或竖直翻转、镜像、旋转、变形、透

视等操作。

(3) 具有不同的效果处理功能，包括可调亮度/对比度、去噪声、模糊、锐化等，还包括其他一些特技。

(4) 具有一定的绘图功能。

不同的图像处理软件其基本功能是类似的，不同点主要在于功能的多少、实现功能的算法，以及窗口界面的使用等。

1. Adobe Photoshop

Photoshop 是由美国 Adobe 公司于 1990 年推出的一个集图像扫描、编辑修改、图像制作、广告创意、图像合成、图像输入/输出于一体的专业图像处理软件。Adobe Photoshop 为美术设计人员提供了无限的创意空间，可以从一个空白的画面或从一幅现成的图像开始，通过各种绘图工具的配合使用及图像调整方式的组合，在图像中任意调整颜色、亮度/对比度、透明度，甚至轮廓；通过几十种特殊滤镜的处理，为作品增添变幻无穷的魅力。

大家所熟悉的 Photoshop 7.0 的下一代产品被命名为 Photoshop CS（详见实验 T2.2），CS 的意思是 Creative Suit。2003 年年底发布了 Photoshop CS，2005 年发布了 Photoshop CS II，最新版本是 Photoshop CS 4，如图 3.68 所示。

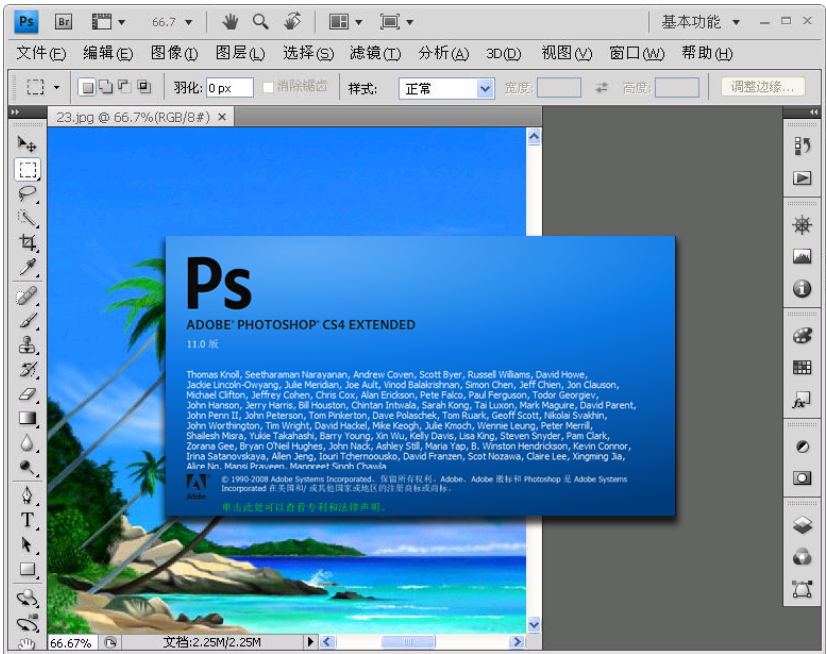


图 3.68 Adobe Photoshop CS 4 界面

2. Adobe Illustrator

作为 Adobe 公司图像处理软件 Photoshop 的姐妹软件，Illustrator 是出版、多媒体和在线图像的工业标准矢量绘图软件。Core 公司的著名绘图软件 CorelDraw 和 Macromedia 公司的 Freehand 也属于矢量绘图软件。Illustrator 是目前市面上最强势、用途最广的矢量绘图软件，它几乎可以与 QuarkXpress, PageMaker, InDesign, Photoshop, Dreamweaver, Flash 等软件

进行最完美的整合，这使得 Illustrator 能够应用于横跨平面、网页与多媒体设计的多种场合。目前的最新版本是 Illustrator CS4，如图 3.69 所示。



图 3.69 Adobe Illustrator CS4 界面

3. Matlab

Matlab（图 3.70）是 MathWorks 公司开发的众多数学软件中最成功的一款，它在发展过程中不断拓宽应用范围，不断增强功能，曾于 1997 年被评为美国十大科技成果之一。Matlab 直接面向科学与工程计算，语言风格接近数学形式，因此比其他高级语言更加接近手工书写计算公式的思维方式，编写简单，效率较高。

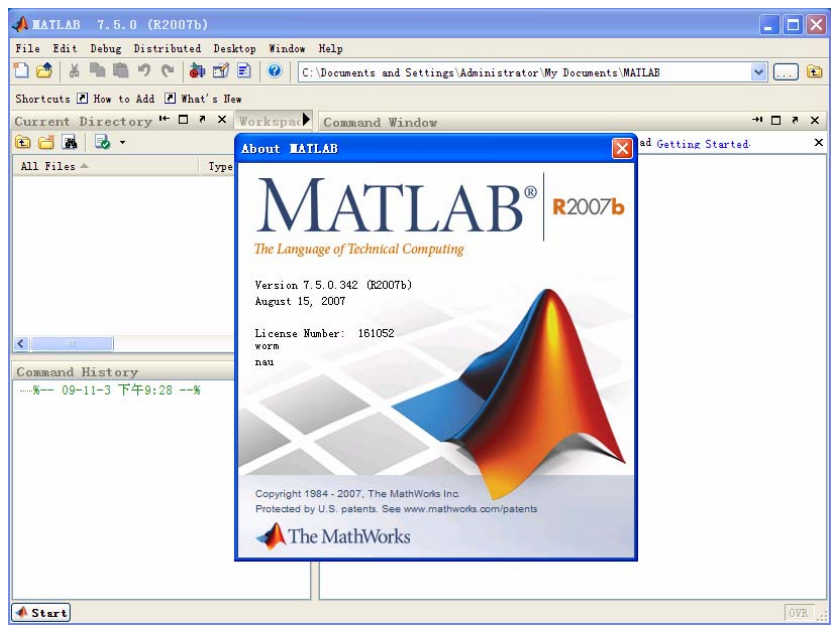


图 3.70 Matlab 界面

Matlab 的图形和图像操作功能很强。在图形方面，它有一系列绘图函数，适合于线性坐标、对数坐标和极坐标等不同坐标显示，此外还能够绘制其他特殊的图形，如切片图、轮廓图、瀑布图、网眼图等。对于图形中的各元素属性的控制也很容易，如颜色、线型、标注等。

在图像方面，它能够显示和生成许多常见格式的图像，甚至是多波段图像。可以对灰度和彩色图像进行所需的各种操作，包括滤波、块操作、几何操作、图像变换、图像分析与增强等。所有这些都是其他高级编程语言的效率所不及的。

Matlab 的优势和特点已经使得它日益成为一种必备的技能，在设计单位和工业部门，Matlab 已经成为研究和解决各种具体工程问题的一种标准软件。

习 题 3

1. 填空题

- (1) 在色彩的三要素中，_____反映颜色的种类，是决定颜色的基本特性。
- (2) 我们在计算机屏幕显示彩色图像时采用的是 RGB 模型，而在打印时一般需要转换为_____模型。
- (3) 图像分辨率与显示分辨率是两个不同的概念。_____是确定组成一幅图像的像素数目的，而_____是确定显示图像的区域大小的。
- (4) GIF 图像格式主要用于图像文件的网络传输，GIF 文件不支持 24 位真彩色图像，最多只能存储_____色的彩色图像或灰度图像。
- (5) 在彩色图像输出设备中，_____打印机已占绝对优势。
- (6) 图像增强和复原的目的是为了提高图像的质量，如去除噪声，提高图像的清晰度等。_____不考虑图像降质的原因，只突出图像中所感兴趣的部分。
- (7) _____技术实际上是计算机认字，它通过扫描和摄像等光学输入方式获取纸张上的文字图像信息，利用各种模式识别算法分析文字形态特征，判断出文字的标准码，并按通用格式存储在文本文件中。
- (8) 指纹识别和人脸识别都属于_____特征识别技术。

2. 简答题

- (1) 图像数字化过程的基本步骤是什么？
- (2) 什么是色彩的三要素？
- (3) 显示分辨率与图像分辨率有什么不同？有一幅分辨率为 320×240 的彩色图像，在显示器分辨率为 640×480 的屏幕上显示，这时图像在屏幕上的显示面积占整个屏幕的多少？如果有一幅分辨率为 1280×960 的彩色图像，显示器的分辨率为 640×480 ，那么在屏幕上只能看到整幅图像的多少？
- (4) 什么是 RGB 模型？为什么将 RGB 色彩的产生方式称为加色法？
- (5) 什么是 CMY 模型？为什么将 CMY 色彩的产生方式称为减色法？
- (6) 像素深度反映了构成图像的颜色总数目，某图像的像素深度为 16 位，则可以显示的颜色数目是多少？
- (7) 什么是真彩色、伪彩色、直接色？

- (8) 试计算 1280×1024 的真彩色图像在未压缩前的数据量。
- (9) 位图和矢量图有什么不一样？各有什么优缺点？
- (10) 常见的数字图像文件格式有哪些？
- (11) 数字图像处理主要研究的内容有哪些？
- (12) 一个 OCR 系统由哪几部分组成？
- (13) 图像处理软件 Adobe Photoshop 通道和图层的概念是什么？

第4章 数字视频

视觉是人类感知外部世界的一个最重要的途径，而计算机视频技术是把我们带到接近真实世界的最强有力的工具。在多媒体技术中，视频信息的获取及处理无疑占有举足轻重的地位，视频处理技术在目前以至将来都是多媒体应用的一个核心技术。

4.1 视频基本概念

人类接受信息的 70% 来自视觉，其中活动图像是信息量最丰富、直观、生动、具体的一种承载信息的媒体。视频（Video）就其本质而言，实际上就是其内容随时间变化的一组动态图像（25 或 30fps），所以视频又称运动图像或活动图像。视频信号具有内容随时间而变化的特性，并有与画面动作同步的声音（伴音）。

图像与视频是两个既有联系又有区别的概念：静止的图片称为图像（Image），运动的图像称为视频（Video）。此外，两者的信源方式不同，图像的输入要靠扫描仪、数字照相机等设备，而视频的输入只能是电视接收机、摄像机、录像机、影碟机，以及可以输入连续图像信号的设备。

按照处理方式的不同，视频分为模拟视频和数字视频。

4.1.1 模拟视频

电视系统的发展前景是数字彩色电视，数字视频的基础是模拟视频，而彩色电视又是在黑白电视的基础上发展起来的。

1. 黑白电视信号

（1）电视的原理

电视是采用动画的视觉原理构造而成的。电视的基本工作原理为顺序扫描和传输图像信号，然后在接收端同步再现。

电视摄像机的作用就是将视频图像转换为电信号。任何时刻，电信号只有 1 个值，即是一维的。但视频图像通常是二维的，将二维视频图像转换为一维电信号是通过光栅扫描实现的。扫描方式主要有逐行扫描和隔行扫描两种。

逐行扫描有以下优点：图像垂直清晰度高，空间处理效果好，有利于电视转换和制式转换，能改善视频压缩效率等。其缺点是码率高，行扫描频率增高，硬件难度加大。

目前的电视系统大都采用隔行扫描，如图 4.1 所示，因为隔行扫描能节省频带，且硬件实现简单。但逐行扫描能获得更好的图像质量和更高的清晰度，不过这是以增加带宽和成本为代价的。隔行扫描行的集合称为场，因此，一帧由两个场组成。一帧电视信号称为一个全电视信号，它由奇数场行信号和偶数场行信号，以及同步信号顺序构成。行信号也就是传输的图像行信号。同步信号是控制信号，所谓同步是指摄像端（发送端）的行、场扫描步调与

显像端（接收端）的扫描步调完全一致，即要求同频率、同相位才能得到一幅稳定的画面。

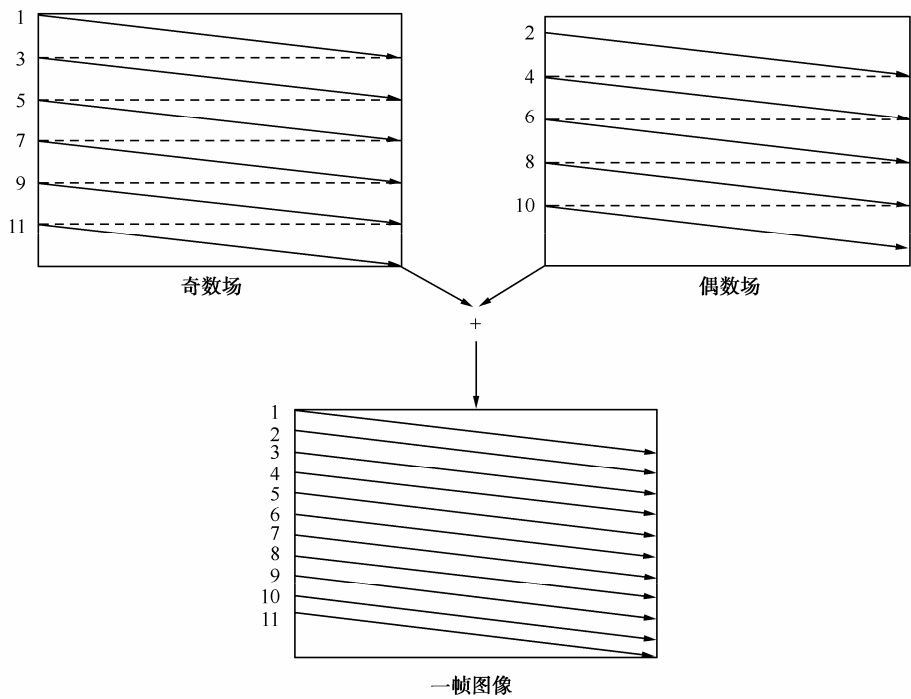


图 4.1 隔行扫描一帧信号的合成过程

(2) 场频、帧频、行频

场频：每秒钟扫描的场数。

帧频：每秒钟扫描的帧数。

行频：每秒钟扫描的行数。

我国黑白电视信号采用每秒 25 帧的标准，每帧图像信号的总行数为 625，分两场扫描。也就是电视信号帧频为 25Hz, 帧周期为 40ms; 场频为 50Hz, 场周期为 20ms; 行频为 15 625Hz, 行周期为 64μs。每场扫描行数为 625/2=312.5 行，其中的前 22 行和后 3 行作为场的回扫和场同步用，不传送图像，因此每场传送图像的行数为 287.5，每帧实际传送图像为 575 行。

(3) 伴音

音频信号的频率范围一般为 20Hz~20kHz，其频带比图像信号窄得多。电视的伴音要求与图像同步，而且不能混叠。因此一般把伴音信号放置在图像频带以外，我国电视信号的声音载波为 6.5MHz，伴音质量为单声道调频广播。

2. 彩色电视信号

1) 彩色与黑白电视信号的兼容

黑白电视只传送一个反映景物亮度的电信号，而彩色电视除了传送亮度信号以外还要传送色度信号。黑白与彩色电视的兼容是指电视台发射一种彩色电视信号，黑白和彩色电视机都能正常工作。为此要将彩色视频信号从 RGB 颜色空间转换到其他颜色空间（如 YIQ, YUV），使亮度信号和色度信号分开传送，使黑白电视和彩色电视能够分别重现黑白和彩色图像。

(1) YUV 模型

在 PAL 彩色电视制式（参见下文“彩色电视的制式”）中采用 YUV 模型来表示彩色图像。其中，Y 表示亮度；U 代表蓝色色差（就是蓝色信号与亮度信号之间的差值，即 $B-Y$ ）；V 代表红色色差（ $R-Y$ ）。U，V 色差信号是构成彩色的两个分量。与此类似，在 NTSC 彩色电视制式中使用的 YIQ 模型，其中，Y 表示亮度；I，Q 是两个彩色分量。YUV 表示法的重要性是它的亮度信号（Y）和色度信号（U，V）是相互独立的，也就是 Y 信号分量构成的黑白灰度图与用 U，V 信号构成的另外两幅单色图是相互独立的，如图 4.2 所示。由于 Y，U，V 是独立的，所以可以对这些单色图像分别进行编码。采用 YUV 模型的优点之一是亮度信号和色差信号是分离的，容易使彩色电视系统与只对亮度敏感的黑白电视机亮度信号兼容。



图 4.2 PAL 彩色电视制式中采用 YUV 模型表示彩色图像

(2) YUV 与 RGB 彩色空间变换

由于所有的显示器都采用 RGB 模型来显示色彩，这就要求在显示每个像素之前，需要把 YUV 彩色分量值转换成 RGB 值。这种转换需要花费一定的计算时间，这是一个在设计软、硬件视频处理系统时要综合考虑的问题。

在考虑人的视觉系统和显示器的非线性特性之后，RGB 和 YUV 的对应关系可以近似地用下面的方程式表示：

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ U &= 0.492(B - Y) = -0.147R - 0.289G + 0.436B \\ V &= 0.877(R - Y) = 0.615R - 0.515G - 0.100B \\ R &= Y + 1.140V \\ G &= Y - 0.395U - 0.581V \\ B &= Y + 2.032U \end{aligned}$$

或者可以写成矩阵的形式：

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$\begin{matrix} R \\ G \\ B \end{matrix} = \begin{matrix} R \\ G \\ B \end{matrix} - \begin{matrix} 0 \\ 0.395 \\ 2.032 \end{matrix} - \begin{matrix} 1.140 \\ 0.581 \\ 0 \end{matrix} \begin{matrix} R \\ G \\ B \end{matrix}$$

注意，这里 R，G，B 的取值范围为[0, 1]；Y 的取值范围为[0, 1]；U 的取值范围为[－0.436；0.436]；V 的取值范围为[－0.615，0.615]。

2) 彩色电视的制式

电视信号的标准也称为电视的制式，目前各国的电视制式不尽相同。目前世界上常用的电视制式有中国、欧洲使用的 PAL 制，美国、日本使用的 NTSC 制及法国、俄罗斯等国所使用的 SECAM 制。这 3 种制式的主要技术指标见表 4.1。

表 4.1 3 种彩色电视制式的主要技术指标

TV 制式	NTSC	PAL	SECAM
帧频（Hz）	30	25	25
行/帧	525	625	625
亮度带宽（MHz）	4.2	6.0	6.0
色度带宽（MHz）	1.3（I），0.6（Q）	1.3（U），1.3（V）	>1.0（U），>1.0（V）
彩色幅载波（MHz）	3.58	4.43	4.25
声音载波（MHz）	4.5	6.5	6.5

(1) NTSC 制

NTSC（National Television Standard Committee）是美国国家电视系统委员会在 1953 年制定的一种兼容的彩色电视制式，是目前常用的视频标准，美国、加拿大等大部分西半球国家，以及中国台湾地区，日本、韩国、菲律宾等均采用这种制式。

NTSC 制采用 YIQ 颜色空间。Y 为亮度信号，色度的 I 分量带宽为 1.3MHz，Q 分量可记录人眼最不敏感的颜色，故可进一步压缩 Q 分量的带宽到 0.6MHz。有的国家（如日本）伴音采用立体声，这样不仅可以传播立体声伴音，还可以实现双语电视节目。双语也就是对应于同一电视画面，两个声道分别广播两种语言，如一个声道传播日语，另一个声道传播英语，两个声道可以互相切换和独立广播。

(2) PAL 制

PAL（Phase Alternate Lock）是联邦德国于 1962 年制定的一种彩色电视制式，它克服了 NTSC 制相位敏感造成色彩失真的缺点。西德、英国等一些西欧国家，新加坡，中国大陆及中国香港地区，澳大利亚、新西兰等国家采用这种制式。PAL 制式中根据不同的参数细节，又可以进一步划分为 G，I，D 等制式，其中 PAL—D 制是我国大陆采用的制式。

(3) SECAM 制

SECAM（Sequential Color and Memory）意为顺序传送彩色信号与存储恢复彩色信号制，是由法国在 1956 年提出的，1966 年制定的一种新的彩色电视制式。它也克服了 NTSC 制式相位失真的缺点，但采用时间分隔法来传送两个色差信号。使用 SECAM 制的国家主要集中在法国、东欧和中东一带。

为了接收和处理不同制式的电视信号，因此也就发展了不同制式的电视接收机和录像机。不同制式的电视机只能接收和处理其对应制式的电视信号。也有多制式或全制式的电视

机，这为处理和转换不同制式的电视信号提供了极大的方便。全制式电视机可在各国各地区使用，而多制式电视机一般为指定范围的国家生产。

3) 彩色电视的信号类型

电视频道传送的电视信号主要包括亮度信号、色度信号、复合同步信号和伴音信号，这些信号可通过频率域，或者时间域相互分离出来。电视接收机能够将所接收到的高频电视信号还原成视频信号和低频伴音信号，并能够在其荧光屏上重现图像，在其扬声器上重现伴音，如图 4.3 所示。

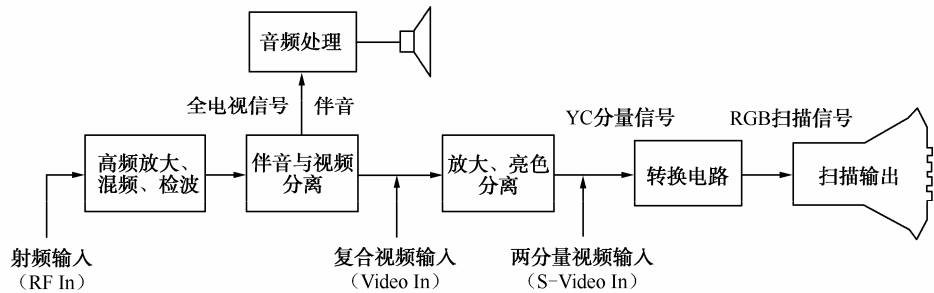


图 4.3 电视机的工作原理

根据不同的信号源，电视接收机的输入、输出信号有如下 3 种类型。

(1) 高频或射频信号

为了能够在空中传播电视信号，必须把视频全电视信号调制成高频或射频 (Radio Frequency, RF) 信号，每个信号占用一个频道，这样才能在空中同时传播多路电视节目而不会导致混乱。PAL 制每个频道占用 8MHz 的带宽；NTSC 制每个频道的带宽为 4MHz。有线电视 CATV (Cable Television) 的工作方式类似，只是它通过电缆而不是通过空中传播电视信号。

电视机在接收到某一频道的高频信号后，要把全电视信号从高频信号中解调出来，才能在屏幕上重现视频图像。

(2) 复合视频 (Composite Video) 信号

为便于电视信号远距离传输，必须把 3 个分量信号及同步信号复合成一个信号，然后再进行传输。复合视频信号定义为包括亮度和色度的单路模拟信号，也就是从全电视信号中分离出伴音后的视频信号，这时的色度信号插在亮度信号的高端。这种信号一般可通过电缆输入/输出到家用录像机上，其信号带宽较窄。

复合视频接口也叫 AV 端子或者 Video 端子 (如图 4.4 所示)，是目前最普遍的一种视频接口，几乎所有的电视机、影碟机类产品都有这个接口。



图 4.4 复合视频接口

AV 端子是声、画分离的视频端子，一般由 3 个独立的 RCA 接口组成。其中的 V 接口连接混合视频信号，为黄色插口，用于复合视频输入和输出（Video In, Video Out），即可以直接输入和输出解调后的视频信号。由于视频信号中已不包含伴音，故一般与视频输入、输出端口配套的还有音频输入、输出端口，以便同步传输伴音。L 接口连接左声道声音信号，为白色插口；R 接口连接右声道声音信号，为红色插口。复合视频信号是一种混合视频信号，没有经过 RF 射频信号那些调制、放大、检波、解调等过程，信号保真度相对较好。不过由于它是模拟接口，当用于数字显示设备时，需要一个模拟转数字（A/D）的过程，会损失不少信噪比，所以一般数字显示设备不建议使用。

（3）分量视频（Component Video）信号与 S-Video

为保证视频信号质量，近距离时可用分量视频信号（Component Video Signal）传输，分量信号是指每个基色分量（R, G, B 或 Y, U, V）作为独立的电视信号传输。计算机输出的 VGA 视频信号，即为分量形式的视频信号。

如将两个单独的色差信号 U 和 V 合为一个色度信号 C（Chroma），再加上单独的亮度信号 Y，即构成 Y/C 信号，也就是 S-Video。S-Video 是一种两分量的视频信号，它把亮度和色度信号（U, V）分成两路独立的模拟信号，用两路导线分别传输并可以分别记录在模拟磁带的两路磁轨上。这种信号不仅其亮度和色度都具有较宽的带宽，而且由于亮度和色度分开传输，可以减少其互相干扰。与复合视频信号相比，S-Video 可以更好地重现色彩。S-Video 端口如图 4.5 所示。



图 4.5 S-Video 端口

两分量视频可来自于高档摄像机，它采用两分量视频的方式记录和传输视频信号。其他如高档录像机的输出也可按分量视频的格式，其清晰度比从家用录像机获得的电视节目的清晰度要高得多。

复合视频信号和分量视频信号的示意图如图 4.6 所示。

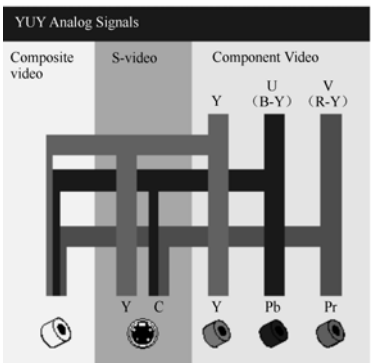


图 4.6 复合视频信号和分量视频信号示意图

4.1.2 数字视频

计算机只能处理数字视频 (Digital Video, DV)，所以，要使计算机能够对视频进行处理，必须把模拟视频信号转换成数字视频。这个过程称为视频的数字化过程（包括采样、量化和编码）。

数字视频克服了模拟视频的局限性，它的主要优点如下。

(1) 适合于网络应用

在网络环境中，视频信息可以很方便地实现资源的共享，通过网络线、光纤，数字信号可以很方便地从资源中心传到办公室和家中。视频数字信号可以长距离传输而不会产生任何不良影响，而模拟信号在传输过程中会有信号损失。

(2) 再现性好

模拟信号由于是连续变化的，所以不管复制时采用的精确度多高，失真总是不可避免的，经过多次复制以后，误差会很大。数字视频可以不失真地进行无限次复制，其抗干扰能力是模拟图像无法比拟的。它不会因存储、传输和复制而产生图像质量的退化，从而能够准确地再现图像。

(3) 便于计算机编辑处理

模拟信号只能简单调整亮度、对比度和颜色等，极大地限制了处理手段和应用范围。而数字视频信号可以传送到计算机内进行存储、处理，很容易进行创造性的编辑与合成，并进行动态交互。

数字视频的缺陷是处理速度慢，所需的数据存储空间大，从而使数字图像的处理成本增高。通过对数字视频的压缩，这样可以节省大量的存储空间，光盘技术的应用也使得大量视频信息的存储成为可能。

如今，数字视频的应用已经非常广泛，并带来一个全新的应用局面。包括数字电视、VCD、DVD、数字式便携摄像机等在内的各种应用均需要采用数字视频。

4.2 视频的数字化过程

要让计算机处理视频信息，首先要解决的是视频数字化的问题。视频数字化是将模拟视频信号经模数转换和彩色空间变换转为计算机可处理的数字信号，与音频、图像信号数字化过程类似，计算机也要对输入的模拟视频信号进行采样与量化，并经编码使其变成数字化图像。

早在 20 世纪 80 年代初，国际无线电咨询委员会 CCIR (International Radio Consultative Committee) 就制定了彩色电视图像数字化标准，称为 CCIR 601 标准，现改为 ITU—R BT.601 标准。该标准规定了彩色电视图像转换成数字图像时使用的采样频率，RGB 和 YCbCr 两个彩色空间之间的转换关系等。

4.2.1 视频信号的采样

1. 彩色空间之间的转换

YCbCr 模型是 YUV 模型的数字化表示，Y 代表亮度信号，Cb, Cr 代表亮度信号。在数

字域而不是模拟域中 RGB 和 YCbCr 两个彩色空间之间的转换关系用下式表示:

$$Y=0.299R+0.587G+0.114B$$

$$Cb=-0.1687R-0.3313G+0.5B+128$$

$$Cr=0.5R-0.4187G-0.0813B+128$$

其中 R, G, B 取值范围为[0, 255]中的整数, Y, Cb, Cr 取值范围也为[0, 255]中的整数。

2. 采样格式

对彩色电视图像进行采样时, 可以采用两种采样方法。一种是使用相同的采样频率对图像的亮度信号和色差信号进行采样, 另一种是对亮度信号和色差信号分别采用不同的采样频率进行采样。如果对色差信号使用的采样频率比对亮度信号使用的采样频率低, 这种采样就称为图像子采样 (Subsampling)。

图像子采样在数字图像压缩技术中得到广泛的应用。可以说, 在彩色图像压缩技术中, 最简便的图像压缩技术恐怕就要算图像子采样了。这种压缩方法的基本根据是人的视觉系统所具有的两条特性, 一是人眼对色差信号的敏感程度比对亮度信号的敏感程度低, 利用这个特性可以把图像中表达颜色的信号去掉一些而使人不易察觉; 二是人眼对图像细节的分辨能力有一定的限度, 利用这个特性可以把图像中的高频信号去掉而使人不易察觉。子采样也就是利用人的视觉系统这两个特性来达到压缩彩色电视信号的目的。

子采样表示为 $4:n:n$ 的格式, 4 代表 ITU—R BT.601 制定的 NTSC, PAL 和 SECAM 视频亮度信号 Y 数字化的标准采样频率 13.5MHz, 后面两个数字分别代表色差信号 Cb 和 Cr 的采样频率。目前使用的子采样格式有如下几种, 如图 4.7 所示, 以视频帧 8 像素×8 像素的宏块 (参见 5.3.1 MPEG—1) 为例。

(1) $4:4:4$ 。这种采样格式不是子采样格式, 它是指在每条扫描线上每 4 个连续的采样点中取 4 个亮度 Y 样本、4 个红色差 Cr 样本和 4 个蓝色差 Cb 样本, 这就相当于每个像素用 3 个样本表示, 如图 4.7 (a) 所示。

(2) $4:2:2$ 。这种子采样格式是指在每条扫描线上每 4 个连续的采样点中取 4 个亮度 Y 样本、2 个红色差 Cr 样本和 2 个蓝色差 Cb 样本, 平均每个像素用 2 个样本表示, 如图 4.7 (b) 所示。当显示像素时, 对于没有 Cb 和 Cr 的 Y 样本, 使用前后相邻的 Cb 和 Cr 样本进行计算得到 Cb 和 Cr 样本。

(3) $4:1:1$ 。这种子采样格式是指在每条扫描线上每 4 个连续的采样点中取 4 个亮度 Y 样本、1 个红色差 Cr 样本和 1 个蓝色差 Cb 样本, 平均每个像素用 1.5 个样本表示, 如图 4.7 (c) 所示。当显示像素时, 对于没有 Cb 和 Cr 的 Y 样本, 使用前后相邻的 Cb 和 Cr 样本进行计算得到该 Y 样本的 Cb 和 Cr 样本。

(4) $4:2:0$ 。这种子采样格式是指在水平和垂直方向上每 2 个连续的采样点中各取 2 个即 4 个亮度 Y 样本、1 个红色差 Cr 样本和 1 个蓝色差 Cb 样本, 平均每个像素用 1.5 个样本表示。虽然 MPEG—2 和 MPEG—1 使用的子采样都是 $4:2:0$, 但它们的含义有所不同, 图 4.7 (d)、图 4.7 (e), 说明了采样格式为 $4:2:0$ 的 YCbCr 空间样本位置。与 MPEG—1 的 $4:2:0$ 相比, MPEG—2 的子采样在水平方向上没有半个像素的偏移。

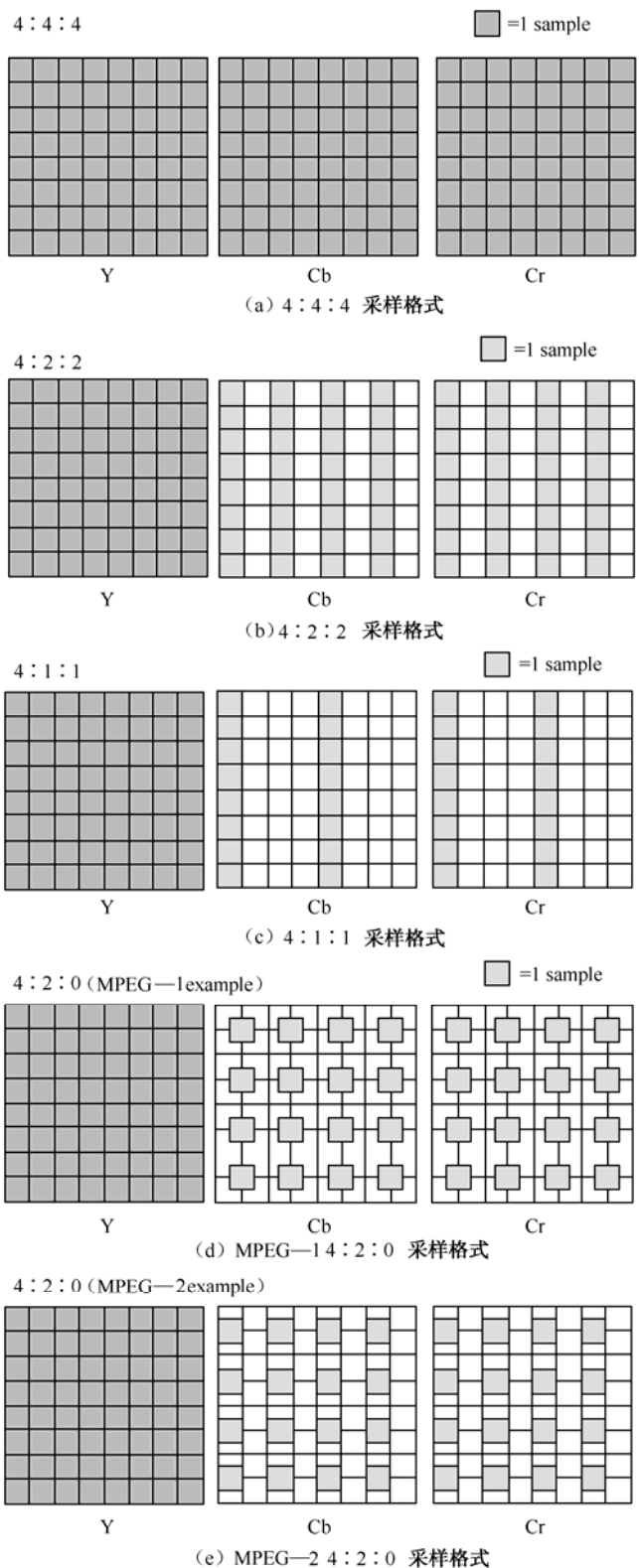


图 4.7 采样格式

3. 采样频率

ITU—R BT.601 用于对隔行扫描电视图像进行数字化,对 NTSC 和 PAL 制彩色电视的采样频率和有效显示分辨率都做了规定。对 PAL 制的亮度信号,每一条扫描行采样 864 个样本;对 NTSC 制的亮度信号,每一条扫描行采样 858 个样本。对所有的制式,每一扫描行的有效样本数均为 720 个,如图 4.8 所示。

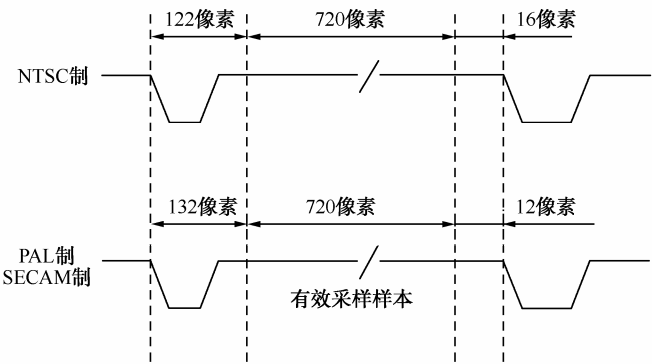


图 4.8 ITU—R BT.601 的亮度采样结构

表 4.2 给出了 ITU—R BT.601 推荐的采样格式、编码参数和采样频率。

表 4.2 ITU—R BT.601 推荐的采样格式、编码参数和采样频率

采样格式	信号格式	采样频率 (MHz)	样本数/扫描行	
			NTSC	PAL
4 : 2 : 2	Y	13.5	858 (720)	864 (720)
	Cb	6.75	429 (360)	432 (360)
	Cr	6.75	429 (360)	432 (360)
4 : 4 : 4	Y	13.5	858 (720)	864 (720)
	Cb	13.5	858 (720)	864 (720)
	Cr	13.5	858 (720)	864 (720)

ITU—R BT.601 推荐使用 4 : 2 : 2 的彩色电视图像采样格式。使用这种采样格式时, Y 用 13.5MHz (对 PAL 制采样频率 $f=625\times25\times N=13.5\text{MHz}$, $N=864$, N 为每一扫描行上的采样数目) 的采样频率, Cb, Cr 用 6.75MHz 的采样频率。采样时, 采样频率信号要与场同步和行同步信号同步。

为了既可用 625 行 (PAL, SECAM) 的电视图像又可用 525 行 (NTSC) 的电视图像, CCITT 规定了公用中分辨率格式 CIF (Common Intermediate Format)、1/4 公用中分辨率格式 QCIF (Quarter-CIF) 和 SQCIF (Sub-Quarter Common Intermediate Format) 格式, 具体格式参数见表 4.3。

表 4.3 CIF, QCIF 和 SQCIF 图像格式参数

参数	CIF		QCIF		SQCIF	
	行数/帧	像素/行	行数/帧	像素/行	行数/帧	像素/行
亮度 (Y)	288	360 (352)	144	180 (176)	96	128
色差 (Cb)	144	180 (176)	72	90 (88)	48	64
色差 (Cr)	144	180	72	90 (88)	48	64

4.2.2 视频信号的量化

采样过程是把模拟信号变成了时间上离散的脉冲信号的过程，量化过程则是进行幅度上的离散化处理的过程。因此在时间轴的任意一点上量化后的信号电平与原模拟信号电平之间在大多数情况下总是存在一定的误差，量化所引入的误差是不可避免的同时也是不可逆的。由于信号的随机性这种误差大小也是随机的，这种表现类似于随机噪声效果，具有相当宽度的频谱，因此我们又把量化误差称为量化噪声。

降低量化误差最直接的方法就是增加量化级数、减小最小量化间隔，但由此带来码率的增加从而要求更大的处理带宽，一般现在的视频信号均采用 8 比特、10 比特，在信号质量要求较高的情况下采用 12 比特量化。

正如模拟音频信号传输过程中采用不均匀量化一样，在视频信号的量化过程中也可以采用不均匀量化方式，即将模拟信号先进行对数变换，其目的是让变化量大的地方变化小，让变化量小的地方变化大。然后，再进行普通的 8 比特量化，经传输后再恢复出来的模拟信号可以通过指数变换予以还原；此时，信号传输的效果类似于 12 比特量化的效果。

4.2.3 视频信号的压缩与编码

采样、量化后的信号转换成数字符号才能进行传输，这一过程称为编码。视频压缩编码的理论基础是信息论。信息压缩就是从时间域、空间域两方面去除冗余信息，将可推知的确定信息去掉。

在通信理论中，编码分为信源编码和信道编码两大类。所谓信源编码是指将信号源中多余的信息除去，形成一个适合用来传输的信号。为了抑制信道噪声对信号的干扰，往往还需要对信号进行再编码，使接收端能够检测或纠正数据在信道传输过程引起的错误，这称为信道编码。

视频编码技术主要包括 MPEG 与 H.26x 标准（参见 5.3~5.4 节），编码技术主要分成帧内编码和帧间编码。前者用于去掉图像的空间冗余信息，后者用于去除图像的时间冗余信息。

4.3 数字视频的获取

数字视频的来源主要有 3 种：一种是利用计算机生成的动画，如把 SWF 或 GIF 动画格式转换成 AVI 等视频格式；另一种是把静态图像或图形文件序列组合成视频文件序列；最后一种，也是最主要的一种是通过视频采集卡把模拟视频转换成数字视频，并按数字视频文件的格式保存下来。第 3 种方式是本节主要讨论的内容。

4.3.1 数字视频系统的组成

从硬件平台的角度分析，数字视频的获取需要 3 个部分的配合：首先是提供模拟视频输出的设备；然后是可以对模拟视频信号进行采集的设备；最后，是接收和记录编码后的数字视频数据的设备，如图 4.9 所示。

提供模拟视频输出的设备有录像机、电视机、影碟机等；对模拟视频信号进行采集、量化和编码的设备由视频采集卡来完成；最后，由计算机接收和记录编码后的数字视频数据。在这一过程中起主要作用的是视频采集卡，如图 4.10 所示，它不仅提供接口以连接模拟视频设备和计算机，而且具有把模拟信号转换成数字信号的功能。

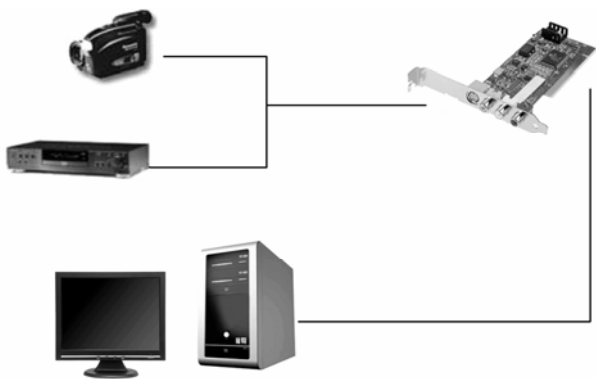


图 4.9 数字视频系统



图 4.10 视频采集（Video Capture）卡

4.3.2 视频采集卡的工作原理

视频采集卡是安装在计算机扩展槽上的一个板卡。它可以汇集多种视频源的信息，如电视、影碟机、录像机和摄像机的视频信息，对被捕捉和采集到的画面进行数字化、冻结、存储、输出及其他处理操作，如编辑、修整、裁剪、按比例绘制、像素显示调整、缩放等。视频采集卡为多媒体视频处理提供了强有力的硬件支持。

视频采集卡的工作原理如图 4.11 所示。视频采集卡一般具有多种视频接口，可接收来自

摄像机、录像机、VCD 机等多种视频信号，通过视频软件可选择所需的视频源。

从彩色摄像机、录像机或其他视频信号源得到的彩色电视信号，经视频接口送入视频采集卡。模/数转换器（ADC）是一个视频解码器，其任务是对视频信号解码和数字化。采用不同的颜色空间可选择不同的视频输入解码器芯片。

经 ADC 解码后得到 YCbCr 信号格式。当以 4：2：2 格式采样时，在每 4 个连续的采样点中取 4 个亮度 Y、2 个色差 Cb、2 个色差 Cr 的样本值，共 8 个样本值。YCbCr 信号经过转换可变成 RGB 信号。

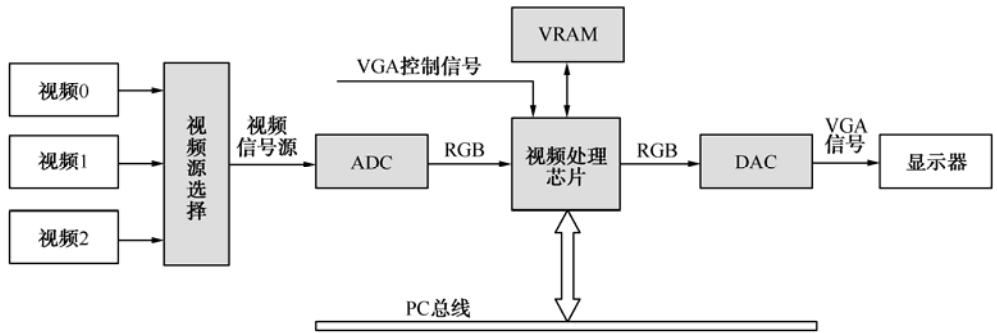


图 4.11 视频采集卡的工作原理框图

RGB 信号然后被送入视频处理芯片，对其进行剪裁、变化等处理。视频处理芯片是用于视频捕获、播放、显示用的专用控制芯片，主要功能可分为 PC 总线读/写、视频输入剪裁、变化比例、与 VGA 信号同步，以及对帧存储器 VRAM（显存）的读/写和刷新控制。

视频信息可实时地存到 VRAM 中，计算机可以通过视频处理器对帧存储器的内容进行读/写操作，帧存储器的视频像素信息读到计算机后，通过编程可以实现各种算法，完成视频图像的编辑与处理。

视频输出的 RGB 信号与显示卡输出的 RGB 信号是完全同步的，可进行叠加，经过 DAC（数/模转换器）转换变成模拟信号，并在显示器的窗口中显示。

由于视频信息量巨大，如果直接存储，会占用大量的存储空间。所以视频采集卡又提供了对视频数字信号的压缩功能，并以压缩的图像文件格式进行存储。当在计算机上播放视频图像时还得经过解压缩过程，使其还原成图像信息才能播放。

4.3.3 视频采集卡的性能指标

根据不同的应用、不同的适用环境和不同的技术指标，目前有多种规格的视频采集卡。可以归纳出以 PC 为硬件环境的视频采集卡的主要功能和技术指标。

1. 接口

视频采集卡的接口包括视频与 PC 的接口和与模拟视频设备的接口。目前 PC 视频采集卡通常采用 32 位的 PCI 总线接口，它插到 PC 主板的扩展槽中，以实现采集卡与 PC 的通信与数据传输。

视频采集卡至少要具有一个复合视频接口（Video In），以便与模拟视频设备相连。高性能的采集卡一般具有一个复合视频接口和一个 S-Video 接口，如图 4.12 所示。一般的采集卡都支持 PAL 和 NTSC 两种电视制式。

视频采集卡如果不具备电视天线接口和音频输入接口，就不能用视频采集卡直接采集电视射频信号，同时也不能直接采集到模拟视频中的伴音信号。要采集伴音，PC 上必须装有声卡，视频采集卡通过 PC 上的声卡获取数字化的伴音并把伴音与采集到的数字视频同步到一起。

一般而言，视频采集卡有单工卡和双工卡两种。单工卡只提供视频输入接口，双工卡还提供输出接口。如果只需在 PC 上编辑数字化视频，单工卡就可以了。若想把数字化编辑过后的影像复制到录像带上，就需要双工卡。

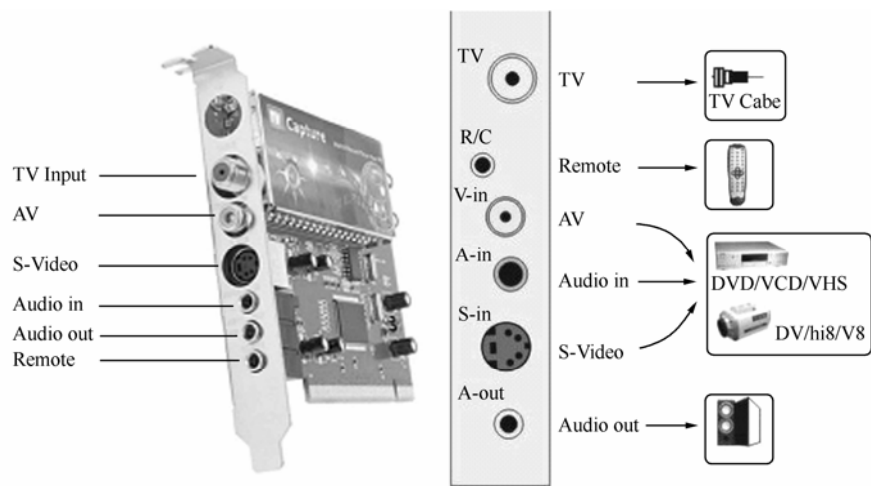


图 4.12 视频采集卡的接口

2. 实时压缩

由于模拟视频输入端可以提供不间断的信息源，视频采集卡要采集模拟视频序列中的每帧图像，并在采集下一帧图像之前把这些数据传入 PC 系统，因此，实现实时采集的关键是每一帧所需的处理时间。如果每帧视频图像的处理时间超过相邻两帧之间的相隔时间，则会出现数据的丢失，即丢帧现象。采集卡都是先把获取的视频序列进行压缩处理，然后再存入硬盘，也就是说，视频序列的获取和压缩是在一起完成的，免除了再次进行压缩处理的不便。不同档次的采集卡具有不同质量的采集压缩性能。

大多数视频采集卡都具备硬件压缩的功能，在采集视频信号时首先在卡上对视频信号进行压缩，然后再通过接口把压缩的视频数据传送到主机上。视频采集卡采用帧内压缩的算法把数字化的视频存储成 AVI 文件，高性能的视频采集卡还能直接把采集到的数字视频数据实时压缩成 MPEG 格式的文件。

视频采集卡一般都采用专业级的专用芯片来处理视频的压缩，同时由于图像采集要处理大量数据，所以对计算机硬件平台的要求也较高。如果用软件的方式进行数据压缩，则对计算机的性能有很高的要求，否则抓取的画面及播放画面的速度便会受到相应的影响。

3. 采集分辨率及帧频

视频采集卡按照其用途可以分为广播级视频采集卡、专业级视频采集卡、民用级视频采集卡。它们的区别主要是采集的图像指标不同。

广播级视频采集卡属高档设备，主要用于电视台制作节目。广播级视频采集卡的最高采集分辨率一般为 720×576 (ITU—R 推荐值)，PAL 制，25fps；或分辨率为 640×480 ，NTSC 制，30fps，最小压缩比一般在 4：1 以内。广播级视频采集卡都带分量输入/输出接口。这一类产品的特点是采集的图像分辨率高，视频信噪比高，缺点是视频文件庞大，每分钟数据量至少为 200MB。

专业级视频采集卡比广播级视频采集卡的性能稍微低一些，分辨率两者是相同的，但压缩比稍微大一些，其最小压缩比一般在 6：1 以内，输入/输出接口为 AV 复合端子与 S 端子。此类产品适用于广告、多媒体节目制作及多媒体软件开发。

民用级的视频采集卡的动态分辨率一般最大为 384×288 ，PAL 制，帧频为 25fps； 320×240 ，30fps，NTSC 制。输入端子为 AV 复合端子与 S 端子，绝大多数不具有视频输出功能。

4. 驱动和应用程序

视频采集卡一般都配有硬件驱动程序以实现 PC 对采集卡的控制和数据通信。根据不同的视频采集卡所要求的操作系统环境，各有不同的驱动程序。只有把采集卡插入了 PC 的主板扩展槽并正确安装了驱动程序以后才能正常工作。采集卡一般都配有采集应用程序以控制和操作采集过程。也有一些通用的采集程序，数字视频编辑软件如 Adobe Premiere 等也带有采集功能，但这些应用软件都必须与采集卡硬件配合使用。也就是说，只有采集卡硬件正常安装和驱动以后才能使用。

4.3.4 视频采集过程

视频采集过程如下。

1. 视频设备的准备

摄像机、录像机等设备都带有复合视频输出端口，有的带有分量视频输出端口。由于视频采集卡提供复合视频输入和分量视频输入，因此只要具有复合视频输出或 S-Video 输出端口的设备都可以为视频采集卡提供视频信号源，把这些输出端口与采集卡相应的视频输入端口相连就可实现信号的连接。

视频的质量在很大程度上取决于模拟视频信号源的质量及视频采集卡的性能，应根据不同的模拟视频信号源分别选择相应的设备。

2. 视频设备与PC的连接

准备好了模拟视频信号源及其相应的设备，剩下的工作就是把模拟视频设备与 PC 上的视频采集卡相连接。模拟设备与视频采集卡的连接包括模拟设备视频输出端口与采集卡视频输入端口的连接，以及模拟设备的音频输出端口与视频采集卡音频输入端口的连接，如图 4.13 所示。

如果采集卡只具有视频输入端口而没有伴音输入端口，要同步采集模拟信号中的伴音，必须使用带声卡的计算机，通过声卡来采集同步伴音。

视频采集卡有两种视频输入接口，要注意它们之间的区别，一种是具有标准复合视频输入的接口 RCA（也称 AV 接口），标准视频信号在输出时要进行编码，将信号压缩后输出，接收时还要进行解码，这样会损失一些信号。还有一种是 S 视频输入接口（S-Video）。由于 S 视频信号不需要进行编码、解码，所以没有信号损失，因此使用 S-Video 接口可以获取更好的图像质量。

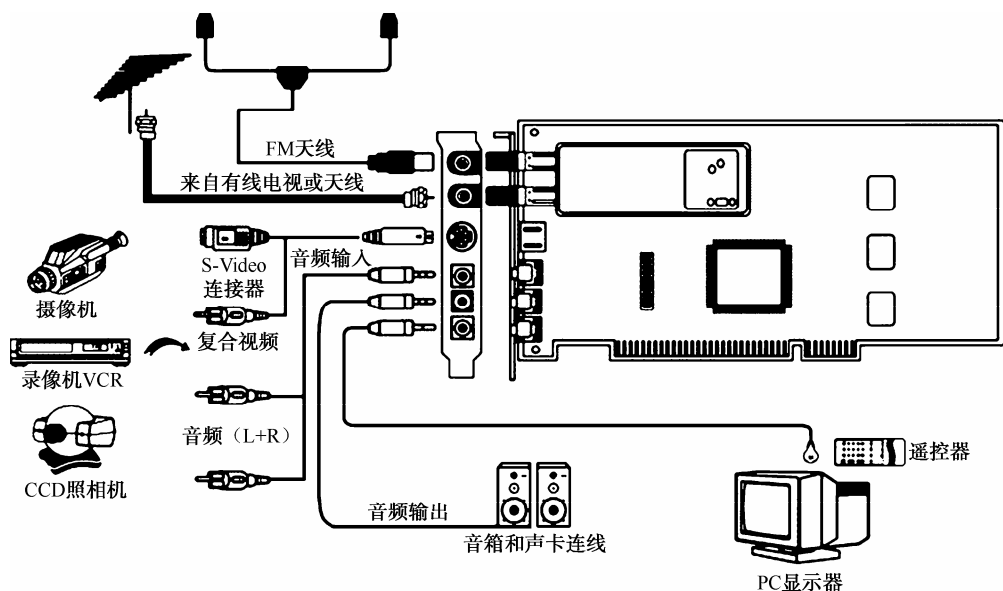


图 4.13 具有多种接口的视频采集卡与视频设备的连接

3. 视频数据的采集

视频设备与 PC 连接好后，可以进行视频数据的采集，其过程主要包括如下几个步骤。

- ① 启动采集程序，设置采集参数（如设置信号源、存放文件格式、存放路径等）。
- ② 启动信号源，然后进行采集。

③ 播放采集的数据，如图 4.14 所示。如果丢帧严重可修改采集参数或进一步优化采集环境，然后重新采集。



图 4.14 播放采集视频数据（所用软件为会声会影 Ulead Video Studio 9.0）

④ 由于信号源是不间断地送往采集卡的视频输入端口的，可根据需要，对采集的原始数据进行简单的编辑。如剪切掉起始和结尾处无用的视频序列，剪切掉中间部分无用的视频序列等，以减少数据所占的硬盘空间。

4.3.5 其他功能的视频设备

1. MPEG卡

MPEG 是能将大量视频信息进行压缩的国际标准。在该标准的支持下，一套 74 分钟的完整录像画面及具有 CD 音质的音频信号，只要一张 CD 光盘即可存储（VCD）。

MPEG 卡实际上分为两类：MPEG 压缩卡和解压卡。

MPEG 压缩卡（如图 4.15 所示）用于将视频影像压缩成 MPEG 的格式。它首先将模拟音视频信号数字化，然后按 MPEG 标准的压缩算法分别对数字音视频信号进行压缩编码，产生一个码率约为 1.5Mbps 的 MPEG 复合音视频码流，最后再转变为.mpg 格式的文件存储在硬盘上。根据所支持的信号输入方式，MPEG 压缩设备可分为专业型和普及型。专业型 MPEG 压缩卡可以支持 YUV、S-Video 和复合视频等多种输入。它们一般还带有数字滤波预处理和专业分量型录像机控制等功能。预处理功能除了能减少视频信号中的噪声外，还可限制视频信号的动态范围，使信号更容易压缩，有效地降低了压缩算法引起的压缩失真，可大大提高图像的主观清晰度。



图 4.15 MPEG 压缩卡

MPEG 解压卡采用硬件方式将压缩后的 VCD 影碟数据解压后进行回放。当计算机将 CD-ROM 内的数据传送到 MPEG 卡上时，通过卡上的 MPEG 解码器，将已压缩的数据进行解压。品质较好的 MPEG 卡可播放每秒 30 帧的视频画面，速度和 NTSC 制式一样。有些 MPEG 卡还提供了视频输出端口（Video Out）和音频输出端口（Audio Out），可以将 VCD 画面播放到大屏幕彩色电视机上或其他录象设备上，具备了视频输出的功能。

2. 电视卡

电视卡（TV Tuner）如图 4.16 所示，其工作原理相当于一台数字式电视机。它首先将从天线接收下来的射频信号变换成视频信号，然后经 A/D 转换器变为数字信号，再经变换电路变为 RGB 模拟信号，最后通过 D/A 转换变为模拟 RGB 信号送到显示器上显示。

因为电视卡采用逐行扫描方式，加上计算机显示点距小，分辨率高，所以整个电视图像看上去清晰稳定，完全可以与电视机媲美。电视卡的硬件部分是电视频道的选台电路，在 MPC 上安装此卡后，允许用户在 MPC 上用遥控器或鼠标进行操作，对电视频道进行选择。



图 4.16 电视卡

不同的电视卡所能选择的频道数量各异。TV Tuner 卡配有声音输出的接口，以供用户连接到音箱或转接到声卡的输入口。除频道选择之外，电视卡还可以进行频道预设、亮度及音量调节、彩色调整等。

电视卡有内置、外置之分，如图 4.16 所示。内置的电视卡插在主板上，外置的可直接接显示器。内置的电视卡必须开机使用，采用软件调频道；外置的打开显示器就可以看电视，采用硬件调频道。内置的需要安装软件播放，可以边看电视边运行其他程序，而且电视窗口大小可以调节，可以对电视节目进行录制；外置的一般看电视时主机就不能使用了，只有全屏看电视，不能录制电视节目。

3. 数字摄像机

将图像信号数字化后存储，这在专业级、广播级的摄录像系统上已应用了相当长时间，但是，它需要付出高昂的代价，因为相应设备的价格很高，一般单位和家庭无法承受。随着数字视频（Digital Video，DV）的标准被国际上 55 个大电子制造公司统一，数字视频正以不太高的价格进入消费领域，数字摄像机也应运而生。

1) 数字摄像机的类型

家用级数字摄像机（如图 4.17 所示）有 MD 摄像机、数字 8mm 摄像机、Mini DV 摄像机之分。MD 摄像机是将信号记录于 MD 光盘上，这种 MD 盘与 MD 随身听所用的 MD 盘原理、尺寸相同，但能连续拍摄的时间较短。

数字 8mm 摄像机是使用 hi8 录像带记录信息的数字摄像机，数字 8mm 摄像机和 Mini DV 摄像机不仅都使用录像带，而且处于同一质量档次，最高拍摄清晰度都达到 500 线水平清晰度以上，外观也非常相似，而数字 8mm 摄像机所用的录像带价格更低。

Mini DV 数字摄像机生产厂家很多，品种型号很多，有更多的选择余地，性能价格比相对较高，体积更小，重量很轻（最轻的 Mini DV 摄像机的重量不足 400g）。



图 4.17 家用级数字摄像机

2) 数字摄像机的特点

数字摄像机是将通过 CCD 转换光信号得到的图像电信号和通过话筒得到的音频电信号，进行模/数转换并压缩处理后送给磁头转换记录，即以信号数字处理为最大特征。就是这一数字化，使它与目前许多家庭广为采用的模拟摄像机相比具有许多优点。

(1) 记录画面质量高

视频图像清晰程度的最基本、最直观的量度是水平清晰度。水平清晰度的线数越多，意味着图像清晰程度越高。由数字摄像机所摄并播放在电视机屏幕上的图像，比人们现在普遍采用的模拟、非广播级摄像机所摄的图像，清晰度要高得多。实际上它可与广播级模拟摄像机所摄的图像质量媲美。目前数字摄像机记录画面的水平清晰度高达 500 线以上（最高 520 线），与早几年广播级的摄像机清晰度水平相当，而家用模拟摄像机记录画面的水平清晰度最高只为 430 线，还有许多只有 250 线。

(2) 记录声音达 CD 水准

DV 摄像机采用两种脉冲调制（PCM）记录方式。一种是取样频率为 48kHz，16 位量化的双声道立体声方式，提供相当于 CD 质量的伴音；另一种是取样频率为 32kHz，12 位量化的 4 声道方式。

(3) 信噪比高

播放录像时在电视画面上出现的雪花斑点是视频噪声。DV 所记录播放的视频信噪比很高，用模拟录像带播放时出现的图像上下颤抖的现象，在以数字方式拍摄记录的录像带上不会出现。

(4) 可拍摄数字照片

数字摄像机也可以像数码相机那样进行数字照相，在 Mini DV 摄像机上有照片拍摄（Photo shot）模式，一旦启用它能够“冻结”、“凝固”一幅幅画面。尽管模拟摄像机有的也能这样做，但用 Mini DV 摄像机所摄的“照片”影像特别清晰，它们不仅可通过电视屏幕显示观看，而且可直接输入计算机进行处理。

(5) 能与计算机进行信息交换

数字摄像机以数字形式记录的图像信号，如能通过接口卡与 PC 相连接，如图 4.18 所示，将信号输入计算机硬盘，就可方便地进行摄像后编辑和多种特技处理。这使得数字摄像机成为多媒体的最佳活动采集源、输入源，而且这种转换无须进行转换与压缩，因此图像几乎没有质量损失和信号丢失，从而便于人们构建数字化的视频编辑系统。

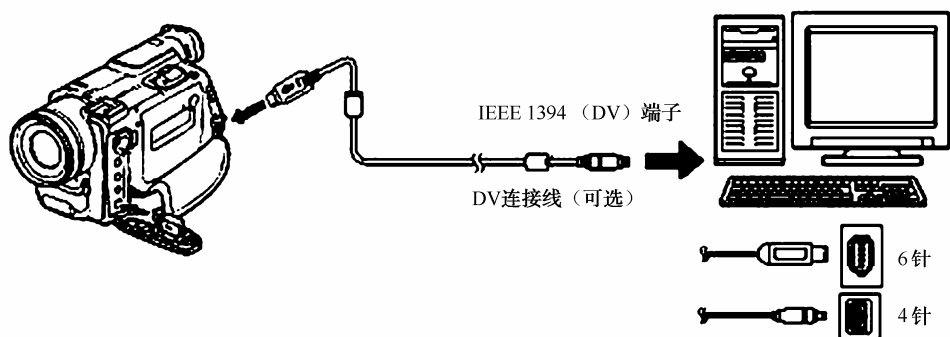


图 4.18 数字摄像机通过 Firewire (IEEE 1394) 接口与计算机相连示意图

4. 数字摄像头

数字摄像头又称为网络摄像机 (Web-Camera) 或计算机摄像机 (PC-Camera)。它是一种新型的多媒体计算机外部设备和网络设备, 人们形象地称之为计算机和网络的“眼睛”。按照工作原理划分, 数字摄像头属于数码影像设备, 其小巧的外形和较好的图像效果等特点, 已经逐渐受到广大用户的欢迎。

最初面世的模拟摄像头必须与视频采集卡一起使用, 才能达到捕捉流畅的动态画面的效果。随着数码影像技术的发展和 USB 接口的普及, 今天的多数数字摄像头都可以通过内部电路直接把图像转换成数字信号传送到计算机上。只要 CPU 处理能力足够快, CCD 捕捉到的图像信号基本上可以达到实时的动态效果。随着计算机 CPU 运算速度的提高, 数字摄像头作为廉价的数码影像产品可以实现一些高档数字设备 (如数码相机、数码摄像机) 的部分功能, 使它具有广阔的市场前景。

网络改变了人们沟通的方式, 用简单的文字交流已不能满足需求, 于是数字摄像头便成为人们在网上进行面对面交流的工具。在局域网甚至 Internet 上, 双方不仅可以通话, 而且可以直接看到对方的图像 (如 QQ 视频聊天), 使网络上的交流更贴近人们的生活习惯, 更符合今天技术发展的潮流。人们更倾向于把数字摄像头看成是一种网络设备, 称之为“网络眼” (Web Eye)。数字摄像头与计算机的连接如图 4.19 所示。

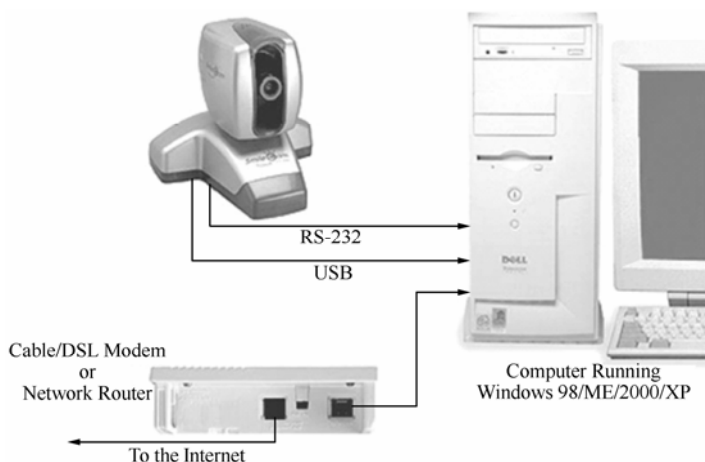


图 4.19 数字摄像头与计算机的连接示意图

数字摄像头是一种依靠软件和硬件配合的多媒体计算机附属设备。其成像使用 CCD 或 CMOS 图像传感器。模拟图像到数字图像的转换等部分与数字摄像机是一样的，只是其光电转换器件分辨率差一些。对数字图像的数据压缩、存储等处理工作，则交给计算机系统去做（可以是硬件卡，但大多数是软件方法）。所以数字摄像头比数码相机、数字摄像机两种数码影像设备都廉价得多。由于数字摄像头主要应用在动态图像捕捉领域，实时捕捉和压缩占用了大量的 CPU 处理时间和内存空间，因此对计算机的硬件处理速度有一定的要求。数字摄像头是否捆绑功能强大的软件，也直接关系到数字摄像头的实际使用效果。大多数数字摄像头都有一个专用的控制程序，以实现最基本的功能，如拍照、摄像、管理影像文件、设置等。

4.4 数字视频的编辑

数字视频的编辑工作主要是依靠软件来完成的，一般的编辑软件都包括素材整理、过场切换和特效、配音、添加字幕等功能。

4.4.1 视频编辑术语

1. 非线性编辑（Non-Linear Editing）

线性编辑（Linear Editing）指的是传统方式下的声像编辑技术。不管是录像带还是录音带，它存储的信息是以时间顺序记录的。当使用者要选取不同的视频素材或某一片段时，需要频繁地倒带，从录像带的一部分找到另外一部分，甚至更换录像带。完成一个编辑工作经常需要反复按顺序找到需要的片段，其过程费时费力，效率较低。

非线性编辑的概念是与“数字化”的概念紧密联系的，是将传统视频编辑系统要完成的工作全部或部分放在计算机上实现的技术，是传统设备同计算机技术相结合的产物。计算机数字化地记录所有视频片段并将它们存储在硬盘上。由于计算机对媒体的交互性，人们可以对存储的数字文件反复地更新或编辑。非线性（Non-Linear）是指访问数字化视、音频信息的方式，非线性表现出数字化信息存取的特点——信息的存取与接受信息的先后顺序无关。非线性技术使得任何片段都可以立即观看并随时任意修改。它利用了计算机软件提供的多种灵活的过渡和特殊效果，可高效地完成原始编辑，如剪辑、切换、动画处理等，再由计算机完成数字视频的生成与计算，并将生成的完整视频回放到视频监视设备或转移到录像带上。由于计算机的交互性及资源的数字化特性突破了传统线性编辑的局限，使视频编辑工作更加随心所欲并富有创造性。

2. 采集/捕捉（Capture）

视频采集这个说法是从模拟视频时代延续下来的。使用模拟视频设备的时候，计算机要想得到视频内容需要使用视频采集卡进行高速的 A/D 转换来完成这个工作。现在随着 DV 的普及，采集数字视频的工作就简单多了，因为不需要进行 A/D 转换，DV 输出的数字信号可以通过 IEEE 1394（FireWire）接口直接保存到计算机中。现在的计算机硬盘速度都足够快，已经不用紧张地“捕捉”了，实际上它们只是在简单地进行复制。

3. 场景/镜头 (Scene)

一个场景也可以称为一个镜头，它是视频作品的基本元素。大多数情况下它指摄像机一次拍摄的一小段内容。对专业摄像人员来说，一个场景大多不会超过十几秒，但是业余用户往往按下拍摄按钮就会停不下来，连续拍摄十几分钟也很常见。所以在编辑过程中经常需要对拍摄的冗长场景进行剪切。视频软件在采集过程中可以通过识别磁带上的时间码来判断独立的场景并进行切分。有的软件在捕捉过程中还可以自动根据镜头内容来分割场景，如果不能的话你也可以手动分割。分割完的场景在编辑的时候可以方便地安排其出现的时间和顺序，而且也有利于插入转场过渡。

4. 字幕/标题 (Title)

字幕/标题泛指在影像中人工加入的所有标志性元素。它可以是文字，也可以是图形、照片、标记等，最常见的应该是字幕。字幕可以像台标一样静止在屏幕一角，也可以做出各种让人眼花缭乱的效果。

5. 转场过渡/切换 (Transition)

两个场景之间如果直接连起来的话，有时会感觉有些突兀。这时使用一个转场效果对两个场景进行过渡就会显得自然很多。最简单的转场过渡就是淡入/淡出效果：前一个场景慢慢暗下去，后一个场景逐渐显示出来。花哨一点的则可以把后一个场景用各种几何分割方式展示出来，再专业一点还能让后面的画面以 3D 方式飞进来等。转场过渡是视频编辑中常用的一个技巧。

6. 特效/滤镜 (Effect/Filter)

动态视频处理中的特效/滤镜和静态图像处理基本相似。这两个名词在视频编辑软件中基本上可以看做同义语，如果非要区分这两个词的含义，或许滤镜更特指亮度、色彩、对比度等方面的调整，而特效更侧重于对影像进行的各种变形和动作效果。多数软件将这些都统称为特效 (Effect)，它们被统一在一个菜单项下，然后又细分为数十种不同类型的特效。

4.4.2 视频编辑软件的基本功能

视频编辑工作曾经需要昂贵的专用设备和大量的专业知识才能完成，而目前大多数家用计算机的性能已经能够胜任这一任务，视频编辑软件的易用性也在变得越来越好。而且这一领域出现了明显的分化：最基本的视频编辑工具能让你组合视频片段并添加简单的转场过渡；中级产品提供了对视频特效和字幕的控制；高端产品则带有更专业的特效、音频和编辑工具。无论你的需求和水平如何，总能从中找出一款适合自己的视频编辑工具。

要想制作出令人赏心悦目的视频节目，首先需要了解视频编辑软件的基本功能。

(1) 基本特效和转场

面向初、中级用户的视频编辑软件会提供直观的特效或转场图标，让你方便地把它们插入视频内容中。

(2) 覆叠

大多数视频编辑软件允许你在视频内容上覆叠一张图片（比如公司的 logo），甚至你还

可以覆盖动态影像来实现画中画。

（3）时间线

几乎所有的视频编辑软件都会提供时间线。在时间线上可以完成切割视频片段、插入转场等操作。

（4）音频

在时间线上还可以放入音频文件，高端产品会提供无限数量的音频轨道及更丰富的音频滤镜。

（5）素材库

素材库中包含了需要在项目中用到的视频、音频、图像等所有素材。

（6）高级特效

高端视频编辑软件提供的特效都带有详细的参数设置，从而让用户进行更精细的控制。

（7）预览窗口

预览窗口是视频编辑软件必备的组件，它可以让你在生成最终结果之前就看到项目的播放效果。

（8）故事板（Storyboard）

故事板是一种显示视频缩略图的视图方式，视频编辑软件往往会提供专门的故事板视图，或在时间线上显示缩略图。

在市场上视频后期编辑软件很多，但大多功能相似。Adobe Premiere（详见实验 3）是 Adobe 公司推出的非常优秀的视频编辑软件。

4.5 数字视频的文件格式

4.5.1 本地视频格式

1. AVI格式

AVI 是音频、视频交错（Audio Video Interleaved）的英文缩写，它是 Microsoft 公司开发的一种符合 RIFF（Resource Interchange File Format，资源交换文件格式）文件规范的数字音频与视频文件格式，原先用于 Microsoft Video for Windows（简称 VFW）环境，现在已被 Windows XP，Linux，OS/2 等多数操作系统直接支持。

AVI 格式允许视频和音频交错在一起同步播放，支持 256 色和 RLE 压缩，但 AVI 文件并未限定压缩标准。因此，AVI 文件格式只是作为控制界面上的标准，不具有兼容性，用不同压缩算法生成的 AVI 文件，必须使用相应的解压缩算法才能播放出来。

AVI 文件目前主要应用在三媒体光盘上，用来保存电影、电视等各种影像信息，有时也出现在 Internet 上，供用户下载、欣赏新影片的精彩片断。

2. DV-AVI格式

DV 的英文全称是 Digital Video，是由索尼、松下、JVC 等多家厂商联合提出的一种家用数字视频格式。目前非常流行的数码摄像机就是使用这种格式记录视频数据的。它可以通过计算机的 IEEE 1394 端口传输视频数据到计算机，也可以将计算机中编辑好的视频数据

回录到数码摄像机中。这种视频格式的文件扩展名一般是 avi，所以也叫 DV-AVI 格式。

3. nAVI格式

nAVI 是 new AVI 的缩写，是一个名为 ShadowRealm 的地下组织发展起来的一种新视频格式。它是由 Microsoft ASF（详见下文网络视频格式）压缩算法的修改而来的（与 AVI 格式没有太大联系），但是又与网络视频格式中的 ASF 视频格式有所区别，它以牺牲原有 ASF 视频文件的“流”特性为代价而通过增加帧率（Frame Rate）来大幅提高 ASF 视频文件的清晰度。概括地说，nAVI 就是一种去掉视频流特性的改良型 ASF 格式，也可以被视为非网络版本的 ASF。

4. MOV格式

QuickTime 是 Apple 公司开发的一种视频文件格式，即 Movie Digital Video 的文件格式，其文件以 MOV 为后缀，相应的视频应用软件为 Apple’s QuickTime for Macintosh。随着大量原本运行在 Macintosh 上的多媒体软件向 PC/Windows 环境的移植，导致了 QuickTime 视频文件的流行。同时 Apple 公司也推出了适用于 PC 的视频应用软件 Apple’s QuickTime for Windows，其界面如图 4.20 所示，因此在 MPC 上也可以播放 MOV 视频文件。



图 4.20 Apple’s QuickTime for Windows 界面

QuickTime 文件格式支持 RLE, JPEG 等集成压缩技术，提供 150 多种视频效果，并配有 200 多种 MIDI 兼容音响和设备的声音装置。新版的 QuickTime 进一步扩展了原有功能，包含了基于 Internet 应用的关键特性，能够通过 Internet 提供实时的数字化信息流、工作流与文件回放功能。

此外,QuickTime 还采用了一种称为 QuickTime VR(简称 QTVR)技术的虚拟现实(Virtual Reality, VR)技术，用户通过鼠标或键盘的交互式控制，可以观察某一地点周围 360° 的景像，或者从空间任何角度观察某一物体。

QuickTime 以其领先的多媒体技术和跨平台特性、较小的存储空间要求、技术细节的独立性及系统的高度开放性，得到业界的广泛认可，目前已成为数字媒体软件技术领域的事实上的工业标准。国际标准化组织 (ISO) 选择 QuickTime 文件格式作为开发 MPEG—4 规范的

统一数字媒体存储格式。

5. MPEG格式——MPEG/MPG/DAT

MPEG 文件格式是运动图像压缩算法的国际标准（详见 5.3 节），它采用有损压缩方法减少运动图像中的冗余信息，已被几乎所有的计算机平台共同支持。MPEG 标准包括 MPEG 视频、MPEG 音频和 MPEG 系统（视频、音频同步）3 个部分，MP3 音频文件就是 MPEG 音频的一个典型应用，而常见的 VCD 属于 MPEG—1 编码，SVCD 和 DVD 都属于 MPEG—2 编码。在网络流行格式中，DivX 编码文件和 XVID 文件，是属于 MPEG—4 编码的。MPEG 压缩标准是针对运动图像而设计的，它主要采用两个基本压缩技术：运动补偿技术（预测编码）实现时间上的压缩；变换域（离散余弦变换 DCT）压缩技术实现空间上的压缩。MPEG 的平均压缩比为 50：1，最高可达 200：1，压缩效率非常高，同时图像和音响的质量也非常好，并且有统一的标准格式，兼容性相当好。

6. DivX格式

这是由 MPEG—4 衍生出的另一种视频编码（压缩）标准，也就是我们通常所说的 DVDrip 格式，它由微软的 MPEG—4 编解码器 MPEG—4 v3（第 3 版）修改而来。其核心技术分为 3 部分：用 MPEG—4 来进行视频压缩，用 MP3 或 AC—3 等压缩音频，同时结合字幕播放软件来外挂字幕。生成的文件包括视频与音频合成生成的 MPEG—4 视频格式文件（AVI 格式），以及相应的外挂字幕文件。其画质直逼 DVD 并且体积只有 DVD 的 1/10 左右。这种编码对机器的要求也不高，所以 DivX 视频编码技术可以说是一种对 DVD 造成威胁最大的新生视频压缩格式。

4.5.2 网络（流式）视频格式

目前，很多视频数据要求通过 Internet 来进行实时传输，视频文件的体积往往比较大，而现有的网络带宽却往往比较“狭窄”，客观因素限制了视频数据的实时传输和实时播放，于是一种新型的流式视频（Streaming Video）格式应运而生了。这种流式视频采用一种“边传边播”的方法，即先从服务器上下载一部分视频文件，形成视频流缓冲区后实时播放，同时继续下载，为接下来的播放做好准备。这种“边传边播”的方法避免了用户必须等待整个文件从 Internet 上全部下载完毕才能观看的缺点。

1. RealVideo格式——RM/RMVB

RealVideo 文件是 RealNetworks 公司开发的一种流式视频文件格式，它包含在 RealNetworks 公司所制定的音频视频压缩规范 RealMedia 中，主要用来在低速率的广域网上实时传输活动视频影像，可以根据网络数据传输速率的不同而采用不同的压缩比率，从而实现影像数据的实时传送和实时播放，使用的播放器是 Real Player，如图 4.21 所示。RealVideo 除了可以以普通的视频文件形式播放之外，还可以与 RealServer 服务器相配合，在数据传输过程中边下载边播放视频影像，而不必像大多数视频文件那样，必须先下载然后才能播放。目前，Internet 上已有不少网站利用 RealVideo 技术进行重大事件的实况转播。



图 4.21 Real Player for Mac OS X 界面

RMVB 影片格式比原先的 RM 多了 VB 两字，在这里 VB 是 VBR（Variable Bit Rate，可变比特率）的缩写。在保证平均采样率的基础上，设定了一般为平均采样率两倍的最大采样率值，在处理较复杂的动态影像时也能得到比较良好的效果，处理一般静止画面时则灵活地转换至较低的采样率，有效地缩减了文件的大小（详见 8.1.2 流媒体文件格式）。

2. MOV 格式

MOV 也可以作为一种流媒体文件格式。QuickTime 能够通过 Internet 提供实时的数字化信息流、工作流与文件回放功能，为了适应这一网络多媒体的应用，QuickTime 为多种流行的浏览器软件提供了相应的 QuickTime Viewer 插件（Plug-in），能够在浏览器中实现多媒体数据的实时回放。该插件的 Fast Start（快速启动）功能，可以令用户几乎能在发出请求的同时便收看到第一帧视频画面。而且，该插件可以在视频数据下载的同时就开始播放视频图像，用户不需要等到全部下载完毕就能进行欣赏。此外，QuickTime 还提供了自动速率选择功能，当用户通过调用插件来播放 QuickTime 多媒体文件时，能够选择不同的连接速率下载并播放影像，当然，不同的速率对应着不同的图像质量。

3. MICROSOFT 流媒体格式——ASF/WMV

ASF（Advanced Streaming Format，高级流格式）是 Microsoft 公司推出的，也是一个在 Internet 上实时传播多媒体的技术标准。它是微软为了和现在的 QuickTime，Real Media 技术竞争而推出的一种视频格式，使用 MPEG—4 的压缩算法，压缩率和图像的质量都很不错。因为 ASF 以“流”格式存在，所以它的图像质量比 VCD 差，但比同是视频“流”格式的 RM 格式要好。这种流式文件的使用与 Windows 操作系统是分不开的，使用的播放器是 Microsoft Media Player，如图 4.22 所示。

WMV（Windows Media Video）也是微软推出的一种采用独立编码方式并且可以直接在网上实时观看视频节目的文件压缩格式，是 ASF 格式的升级和延伸。在同等视频质量下，WMV 格式的体积非常小，因此很适合在网上播放和传输。WMV 格式的主要优点包括：本地或网络回放、可扩充的媒体类型、部件下载、可伸缩的媒体类型、流的优先级化、多语言

支持、环境独立性、丰富的流间关系及扩展性等。



图 4.22 Windows Media Player 界面

4.6 数字视频文件格式转换

随着计算机影像技术的不断发展，各种视频格式不断涌现，本节将简要介绍各种主流视频格式间的转换工具。

4.6.1 全能转换工具

1. WinMPG Video Convert

这是一款非常优秀的全能视频转换工具，软件不仅可以同时将 AVI 格式转换成 MPEG—1/MPEG—2/DVD/VCD/SVCD/DivX 等视频格式，而且还可以将 ASF 和 WMV 等视频格式快速转换成 AVI 及 MPEG—4 格式，如图 4.23 所示。



图 4.23 WinMPG Video Convert 界面

下面以将一个 VCD 文件转换成 DivX 格式的视频文件为例说明软件的使用方法：用户在软件界面上方单击“ALL to DivX”按钮并在接下来弹出的对话框的“Input File”（输入文件）空白框中指定需要转换 VCD 文件的文件存放路径，软件会自动在对应的“Output File”（输出文件）空白框中填入一个相同的转换后文件的输出路径并会在原文件名中加上“.avi”扩展名，如果用户对软件默认的转换文件存放路径不满意可以随时进行更改。另外，用户通过选中“Preview”（预览）选项还可以即时对转换视频进行预览，如图 4.24 所示。进行完所有准

备工作后，单击“Convert”按钮，软件即可马上进入视频格式转换状态。



图 4.24 WinMPG Video Convert 转换并预览界面

2. Zealot All Video Converter

这是一款可以分别将 AVI 转换成 nAVI/MPEG—1/MPEG—2/DVD/VCD/SVCD/WMV/ASF 视频文件和将 MPEG 转换成 AVI/DivX/WMV/ASF 视频文件的视频文件转换工具，如图 4.25 所示。



图 4.25 Zealot All Video Converter 界面

用户可以在工具栏中单击“添加”按钮，将需要进行转换的 AVI、MPEG 或 MPG 文件添加到软件的转换列表中，然后在列表中选中需要转换的视频文件，单击鼠标右键并选择需要转换的视频类型即可。这款工具有 3 个特色：一是支持对被转换文件和转换后的视频文件进行即时预览；二是用户可以在软件的“选项”对话框中对各个具体视频转换类型的视频参数进行定制，如图 4.26 所示，如将一个 AVI 文件转换成 VCD 格式文件或 DVD 格式文件，则可以对 MPEG 类型（MPEG—1 或 MPEG—2）、帧速率和视频宽高比例等进行定制；三是软件支持同时对多个文件进行批量视频转换。

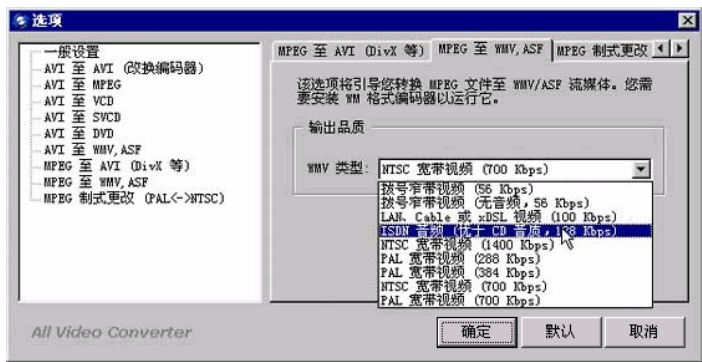


图 4.26 Zealot All Video Converter 转换选项设置

3. AVS Video Converter

这是一款功能特别全面的视频文件转换及编辑工具，如图 4.27 所示。

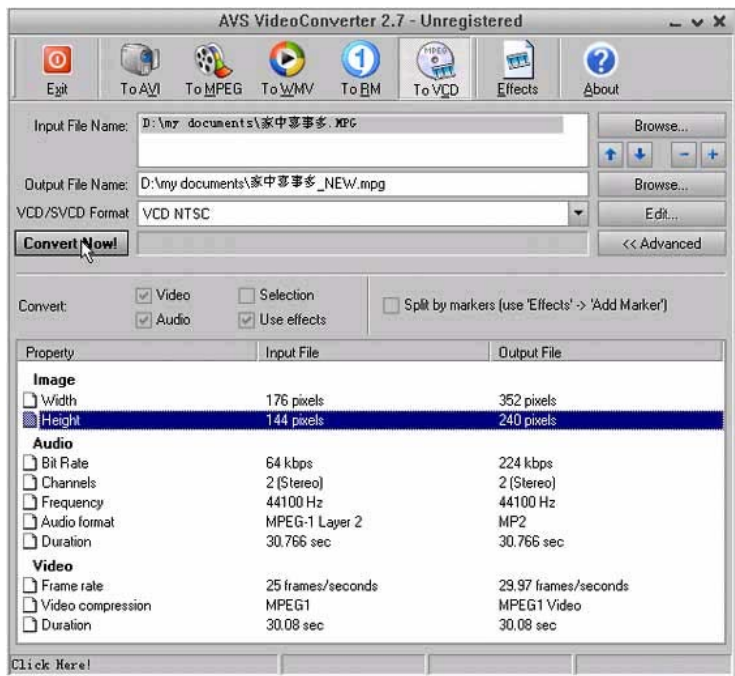


图 4.27 AVS Video Converter 界面

用户在软件的主界面中单击“Browse”（浏览）按钮将需要转换的 AVI/MPG/MPEG—1/MPEG—2/VOB（DVD）/DAT（VCD）/WMV/ASF 等格式的视频文件导入到软件中后，直接在软件的工具栏中单击相应的 To AVI, To MPEG, To WMV, To RM 和 To VCD 按钮软件，即可快速完成对一个视频文件转换任务的定制。如果用户对软件默认的视频文件转换任务参数不太满意的话，则还可以在“Output File Name”（输出文件名）和“Advanced”（高级）选项中对转换文件的存放路径及视频转换参数进行详细定制。在设置完成后，用户在软件界面左上方单击“Convert Now!”（开始转换）按钮，软件即可进行指定视频格式的转换。另外，在视频编辑方面，用户可以同时对多段视频片段进行转换合并使之成为一部完整影片，或者用户也可以对一部完整影片进行多重剪辑并通过各种视频格式的转换使之成为可以满足不同需要的多段不同格式的视频片段。另外，在进行视频文件转换及编辑过程中，软件均支持视频效果的即时预览。

4.6.2 RM相关转换工具

1. 网络多媒体梦工场

这是一款可以直接将 VCD（.dat）、DVD（.vob）、.mpg 和.wav 等视频格式转换成.rm 格式的多功能音视频转换工具(在音频格式转换方面软件支持将 MP3 和 CD 直接转换成 RM 音频格式)，如图 4.28 所示。

用户在该软件的工具栏中单击“视频”按钮并通过软件界面左侧的文件树状目录将需要转换的视频文件导入到转换列表，然后选中需要转换的视频文件，在工具栏中单击“开始”，则软件会自动将指定的视频格式转换成 RM 文件。



图 4.28 网络多媒体梦工场界面

2. RM Converter

这是一款可以将 Real Media 文件 (*.rm, *.rmvb) 格式转换成 AVI, MPEG, VCD, SVCD, DVD, WMV 等格式文件的实用小工具，如图 4.29 所示。

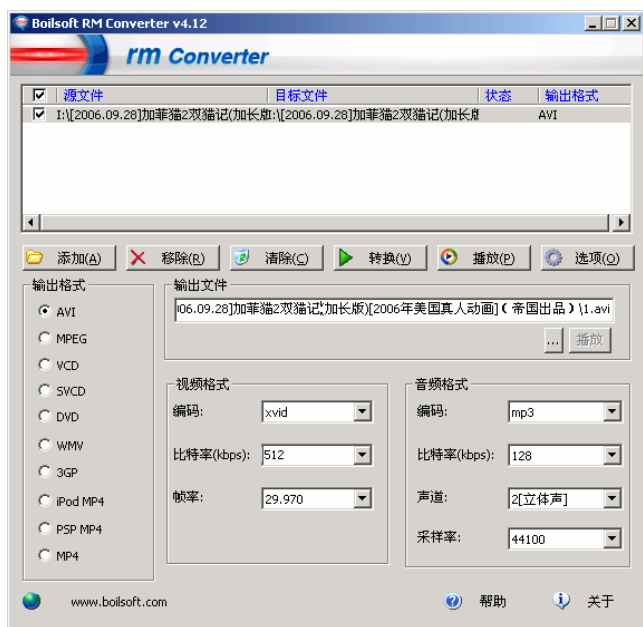


图 4.29 RM Converter 界面

这款工具非常好用，用户只要指定需要转换的 Real Media (*.rm, *.rmvb) 文件，以及转换后文件格式和输出路径，然后再单击“转换”(Convert) 按钮即可完成文件的格式转换。另外，这款工具还可以支持用户对格式转换过程中的转换参数（如比特率、帧率的大小等）进行设定。

4.6.3 DivX相关转换工具

这是一款可以将 AVI/MPEG/WMV/VOB/ASF 等格式的视频文件快速转换成 DivX（扩展名为.avi）格式文件的小工具，如图 4.30 所示。

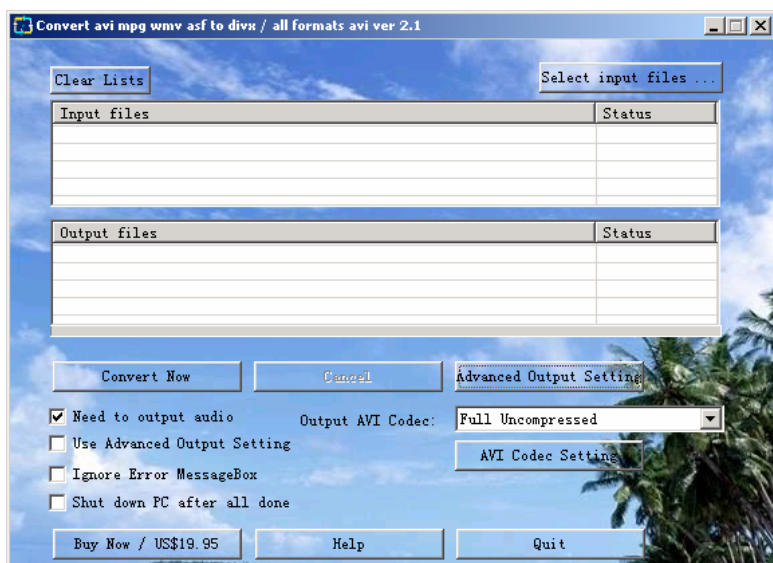


图 4.30 AVI to DivX 界面

将其其他格式的视频转换为 DivX 格式的最大好处是可以使用比较小的文件存储空间存储尽可能多的高质量视频资料。用户只要在软件界面右上方将需要转换的视频文件添加到“Input Files”列表中（软件支持视频文件批量导入），然后修改必要的转换参数选项。完成所有准备工作后，用户单击“Convert Now”按钮即可进行视频转换。

4.6.4 DVD 相关转换工具

这是一款将 DVD 视频格式进行全方位格式转换的超级工具，豪杰在国内是一家具有品牌效应的知名公司，软件品质很有保证，如图 4.31 所示。

这款工具不仅可以将 DVD 向 VCD 进行视频转换（转换过程中可以选择字幕、语言和角度生成 VCD 格式的 MPEG 文件），而且软件还可以支持将 DVD 视频直接转换成 MPEG—4/AVI/RM 等视频格式。在进行视频转换的过程中，用户可以对各种转换格式的转换参数进行详细定制，并且可以对需要转换的视频文件进行预览。



图 4.31 豪杰视频通界面

习 题 4

1. 填空题

- (1) 电视信号的标准也称为电视的制式，各国的电视制式不尽相同。目前世界上常用的电视制式有中国、欧洲使用的_____制，美国、日本使用的_____制及法国、俄罗斯等国所使用的_____制。
- (2) S-Video 是一种_____分量的视频信号。
- (3) _____不仅提供接口以连接模拟视频设备和计算机，而且具有把模拟信号转换成数字信号的功能。

2. 简答题

- (1) 什么是视频？有哪些特点？
- (2) 逐行扫描与隔行扫描各有什么优缺点？
- (3) PAL 制式和 NTSC 制式彩色电视各使用什么颜色模型？计算机图像显示使用什么颜色模型？
- (4) 为什么黑白电视机可以接收彩色电视信号？
- (5) 数字视频常用的采样格式为 4:2:2，它的含义是什么？
- (6) 视频采集卡的工作原理是什么？有哪些性能指标？
- (7) 简述视频采集过程。
- (8) 数字摄像机有什么特点？
- (9) 数字摄像头和数字摄像机的原理有什么不同？
- (10) 如何实现在计算机上看电视？
- (11) 数字视频文件的类型有哪些？
- (12) 常用的视频文件格式转换工具有哪些？

3. 操作题

- (1) 视频采集实验。

利用已有的视频采集系统，根据不同的采集条件分别采集同一段视频素材。

- ① 观察视频采集系统的连接方式，画出其主要接口的连接示意图。
 - ② 记录采集系统中计算机的性能配置。
 - ③ 开启采集系统，预览模拟视频源，选择一段要采集的内容，调整图像效果，确定视频存放路径。
 - ④ 根据不同的采集条件，分别采集并保存该段视频文件，注意采集过程中丢帧的情况。
 - ⑤ 浏览④中按不同条件采集到的数字视频文件，并记录不同的播放效果。
- (2) 使用 **Premiere** 对一段视频进行剪辑，添加字幕、特效及音效的处理。

第5章 多媒体数据压缩

多媒体计算机要处理的信息主要有文字、声音、图形（矢量图、动画）、图像（位图形式的静态图像、动态图像，包括视频）等，而需要处理的图形与图像信息约占总信息量的 85%。所以，对多媒体信息进行压缩的目的是减小存储容量和降低数据传输率，使得现有 PC 在指标与性能方面达到能够处理声音与图像信息的要求，这是建立在多媒体计算机硬件支撑平台所必须具备的功能上的。在这个过程中，声音与图像的信息都需要进行压缩处理。但其中矛盾最为突出和困难的是图像信息压缩，这是因为数字化后的图像信息的数据量非常大，使得存储与处理都十分困难。数据压缩技术的重要作用在图像信息的压缩方面表现得尤为明显。

多年来人们一直致力于图像压缩技术的研究，并产生了很多有用的算法和技术。为了方便这些技术的使用，目前也制定出了一些用于各种应用的图像压缩的国际标准。本章主要介绍图像与视频数据的常用压缩算法及标准。

5.1 数据压缩的基本概念

5.1.1 为什么要数据压缩

信息时代的重要特征是信息的数字化，数字化了的信息带来了“信息爆炸”。多媒体计算机面临声音、图像、视频等多种媒体数字化信息的存储和传输问题。未经压缩的数字化多媒体信息数据量是非常惊人的。

【例 5.1】 计算没有压缩前音频、静态图像和高清晰数字电视视频数据量。

（1）音频数据量

对于音频信息来说，如果按 CD 音质（CD-DA）对原始音频进行不经压缩的数字化，采样频率为 44.1kHz、量化精度为 16 位、双声道立体声，则每分钟的数据量为

$$44.1 \times 10^3 \times 16 \times 2 \times 60 \div 8 \div 2^{20} = 10.1\text{MB}$$

这样，一张 CD-ROM 光盘按 650MB 的容量来计算，只能存放 1 小时的 CD 音乐。

（2）静态图像数据量

目前数码相机主流的分辨率为 500 万像素（2576×1932dpi），如果按 24 位色深来表达，则存储这张图片所需的磁盘空间为

$$5\,000\,000 \times 24 \div 8 \div 2^{20} = 14.3\text{MB}$$

则一张 128MB 的存储卡，只能存储 8 张照片。

（3）高清晰数字电视视频数据量

以计划中的高清晰数字电视视频数据为例，其最高分辨率达 1920×1152（25fp/s）（采用 MPEG—2 的 MP@HL 编码方案），则其每秒视频数据量高达

$$1920 \times 1152 \times 24 \div 8 \div 2^{20} \times 25 = 158.2\text{MB}$$

这样，一张 CD-ROM 光盘按 650MB 的容量来计算，只能存放 4s 的高清晰电视节目。

多媒体信息的数据量过大使得利用计算机对多媒体数据进行处理面临很大的困难,对存储器的存储容量、通信线路的传输速率及计算机的速度都增加了极大的压力。这个问题是多媒体技术发展中的一个非常棘手的瓶颈问题。解决这个问题,单纯用扩大存储器容量、增加通信干线的传输速率的方法是不现实的。数据压缩技术是一个行之有效的方法。通过数据压缩手段,以压缩形式存储和传输信息,既节约了存储空间,又提高了通信干线的传输效率,同时也使计算机实时处理音频、视频信息,以保证播放出高质量的视频、音频节目成为可能。

5.1.2 多媒体数据压缩的发展

1. 经典压缩编码技术

1948年,Oliver提出了第一个编码理论——脉冲编码调制(Pulse Code Modulation, PCM);同年,现代信息论的创始人香农(C.E.Shannon)的经典论文——“通信的数学原理”首次提出并建立了信息率失真函数概念;1959年,香农进一步确立了码率失真理论,以上工作奠定了信息编码的理论基础。主要编码方法有预测编码、变换编码和统计编码,也称为3大经典编码方法。经典编码技术又可以称为“第一代”编码技术。

“第一代”视频(图像)编码技术主要根据信息论原理去除图像的相关性(冗余度),它们能够在中等压缩率的情况下,提供非常好的图像质量。20世纪80年代初期,“第一代”编码技术已经达到了顶峰,这类技术去除图像冗余信息的能力已接近极限。究其原因这是由于这些技术都没有利用图像的结构特点,因此它们也就只能以像素或块作为编码的对象,另外,这些技术在设计编码器时也没有考虑人类视觉系统的特性。

2. “第二代”压缩编码方法

“第一代”编码技术只是以信息论和数字信号处理技术为理论基础,旨在去除图像数据中的线性相关性的一类编码技术。其压缩比不高,大约在10:1左右。为了克服“第一代”视频(图像)编码技术的局限性,Kunt等人于1985年提出了“第二代”视频(图像)编码技术。“第二代”编码技术不局限于信息论的框架,而是充分利用人的视觉生理、心理和图像信源的各种特征,实现从“波形”编码到“模型”编码的转变,以便获得更高压缩比。其压缩比多在30:1至70:1之间,甚至更高。“第二代”编码方法主要有基于分形的编码、基于模型的编码、基于区域分割的编码和基于神经网络的编码等。

随着产业化活动的进一步开展,国际标准化组织于1986年、1998年先后成立了联合图片专家组JPEG和运动图像压缩编码组织MPEG。JPEG专家组主要致力于静态图像的帧内压缩编码标准ISO/IEC 10918的制定;MPEG专家组主要致力于运动图像压缩编码标准的制定。经过专家组不懈的努力,基于第一代压缩编码方法(如预测编码、变换编码、熵编码及运动补偿等)的三种压缩编码国际标准正式编辑出版了:

- JPEG,即静态图像压缩编码标准
- MPEG—1,即数字声像存储压缩编码标准
- MPEG—2,即通用视频图像压缩编码标准

随后,MPEG专家组于1999年2月正式公布了MPEG—4(ISO/IEC 14496) V1.0版本。同年底MPEG—4 V2.0版本亦告完成,且于2000年年初正式成为国际标准。MPEG—4标准将众多的多媒体应用集成于一个完整的框架内,旨在为多媒体通信及应用环境提供标准的

算法及工具，从而建立起一种能被多媒体传输、存储、检索等应用普通采用的统一数据格式，并根据不同的应用需求，现场配置解码器，开放的编码系统也可随时加入新的有效的算法模块。

目前，又推出了专门支持多媒体信息基于内容检索的编码方案 MPEG—7，以及多媒体框架标准 MPEG—21。另外，由 ITU-T 和 MPEG 联合开发的新标准 H.264 是最新的视频编码算法。为了降低码率，获得尽可能更好的图像质量，H.264 标准吸取了 MPEG—4 的长处，具有更高的压缩比、更好的信道适应性，必将在数字视频的通信和存储领域得到越来越广泛的应用，其发展潜力不可限量。

5.1.3 多媒体数据压缩的分类

多媒体数据压缩方法根据不同的依据可产生不同的分类。例如，根据信息有无损失可分为以下几类。

(1) 无损压缩

无损压缩是指使用压缩后的数据进行重构（或称还原，解压缩），重构后的数据与原来的数据完全相同，无损压缩用于要求重构的信号与原始信号完全一致的场合。一个很常见的例子是磁盘文件压缩。根据目前的技术水平，无损压缩算法一般可把普通文件数据压缩到原来的 $1/2 \sim 1/4$ 。一些常用的无损压缩算法有哈夫曼算法和 LZW 压缩算法等。

(2) 有损压缩

有损压缩是指使用压缩后的数据进行重构，重构后的数据与原来的数据有所不同，但不会让人对原始资料表达的信息造成误解。有损压缩适用于重构信号不一定非要和原始信号完全相同的场合。例如，图像和声音的压缩就可以采用有损压缩，因为其中包含的数据往往多于我们的视觉系统和听觉系统所能接收的信息，丢掉一些数据也不至于对声音或者图像所表达的意思产生误解，但可大大提高压缩比。

根据其作用域在空间域或频率域上分为空间方法、变换方法和混合方法。根据是否自适应分为自适应性编码和非自适应性编码。

5.1.4 多媒体数据压缩性能的评价标准

一个好的压缩方法对多媒体信息的存储和传输是至关重要的，衡量一个压缩编码方法优劣的性能指标主要有如下三点：

(1) 压缩比

压缩比是指文件原始大小和经压缩后文件大小之间的比例。作为压缩率的衡量指标，人们普遍希望压缩的倍数越高越好。

(2) 压缩质量

虽然有损压缩可获得较大的压缩比，但压缩比过高，还原后的图像质量就可能降低。人们在追求高压缩比的同时，又希望确保压缩的质量，即压缩完了以后，解压缩的数据和原来的数据最好没有什么差别，没有什么数据损失。

静态图像的压缩标准 JPEG2000 与 JPEG 相比，具有在低比特率的条件下，也能提供相当优质的图像的特点。JPEG2000 与 JPEG 相比不仅压缩比较高，解压缩重建后的图像质量也较好。

【例 5.2】 某风景图片 JPEG2000 与 JPEG 静态图像的压缩质量比较。

某风景图片原图像显示如图 5.1 (a) 所示, 由原图像向 JPEG2000 转换后的图像显示如图 5.1 (b) 所示, 由原图像向 JPEG 转换后的图像显示如图 5.1 (c) 所示。

从图 5.1 中可以看出 JPEG2000 与 JPEG 相比不仅压缩比较高, JPEG2000 的压缩比为 $819\ 894/5047 \approx 162:1$, 而 JPEG 的压缩比为 $819\ 894/8227 \approx 100:1$, 解压缩重建后的图像质量也较好。



(a) 原图像
Windows 位图格式
640×427 (像素)
819 894 B



(b) 由原图像向 JPEG2000 转换
JPEG2000 格式
640×427 (像素)
5047 B



(c) 由原图像向 JPEG 转换
JPEG 格式
640×427 (像素)
8227 B

图 5.1 JPEG2000 与 JPEG 压缩质量比较

(3) 压缩与解压缩的速度

压缩和解压缩的速度是压缩系统的两项单独的性能度量。在有些应用中, 压缩和解压缩都需要实时进行, 压缩和解压缩的算法是一样的, 称为对称压缩, 如视频会议这种实时传递的系统便采用对称压缩技术。发送方将实况视频信号用某种算法加以压缩, 然后通过通信介质进行传输。接收端在收到信号后, 再使用同样的算法按逆运算进行解压缩, 使图像解码后重现出来。在有些应用中, 压缩可以用非实时压缩, 而只要解压缩是实时的, 这种压缩称为非对称压缩, 如多媒体 CD-ROM 的节目制作。从目前开发的压缩技术来看, 一般压缩的计算量比解压缩要大。

压缩的速度不仅与采用的压缩方法有关, 而且与快速算法的计算量有关, 如果在算法上有较大的突破, 无疑将对多媒体的开发与应用产生很大的影响。

5.2 静态图像压缩标准JPEG

JPEG (Joint Photographic Experts Group) 是指由国际标准化组织 (ISO) 和国际电

报电话委员会（CCITT）联合成立的专家组联合制定的一个适用于连续色调、多级灰度、彩色或单色静止图像数据压缩的国际标准。JPEG 方案的问世，在多媒体技术领域产生了巨大的影响，并迅速应用于视频压缩编码国际标准 MPEG 中。JPEG 以其较大的压缩比和很好的压缩效果，对网络多媒体的应用、多媒体系统集成等产生了极其重要的推动作用。

5.2.1 JPEG算法概要

JPEG 专家组开发了两种基本的压缩算法，一种是采用以离散余弦变换（Discrete Cosine Transform, DCT）为基础的有损压缩算法；另一种是采用以预测技术为基础的无损压缩算法。当使用有损压缩算法时，在压缩比为 25：1 的情况下，压缩后还原得到的图像与原始图像相比较，非图像专家很难找出它们之间的区别，因此得到了广泛的应用。例如，在 VCD 和 DVD-Video 电视图像压缩技术中，就使用 JPEG 的有损压缩算法来消除冗余数据。为了在保证图像质量的前提下进一步提高压缩比，1997 年 3 月，JPEG 又开始着手制定用于静态图像压缩的更优秀的方案，该方案采用以小波变换（Wavelet Transform）为主的多分辨率编码方式，并命名为 JPEG2000，且于 1999 年 11 月公布为国际标准，成为图像压缩领域又一项具有划时代意义的技术。

JPEG 压缩是有损压缩，它利用了人的视觉系统的特性，将量化和无损压缩编码结合起来去掉视觉的冗余信息和数据本身的冗余信息。

5.2.2 JPEG算法的主要步骤

JPEG 压缩算法主要分 4 个步骤来实现：颜色模型转换及采样、DCT 变换、量化和编码。解压缩的过程与压缩编码过程正好相反。

1. 颜色模型转换及采样

RGB 色彩模型是我们最常用的表示颜色的方式。JPEG 采用的是 YCbCr 颜色模型（参见 4.2 节）。这里，Y 代表亮度；Cb 和 Cr 则代表色度。JPEG 为什么使用的是 YCbCr 颜色模型，而不是 RGB 颜色模型，其中的一个主要理由是人眼对亮度的改变比对色彩的改变要敏感得多。也就是说，Y 成分的数据是比较重要的。既然 Cb 成分和 Cr 成分的数据比相对不太重要，那么就可以只取部分数据来处理，以增加压缩的比例。

JPEG 通常有两种采样方式：4：1：1 和 4：2：2，它们所代表的意义是 Y，Cb 和 Cr 三个成分的数据采样比例。

想要用 JPEG 基本压缩算法处理全彩色图像，得先把 RGB 颜色模型的图像数据转换为 YCbCr 颜色模型的数据。通过下列计算公式可完成数据转换：

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ Cb &= -0.1687R - 0.3313G + 0.5B + 128 \\ Cr &= 0.5R - 0.4187G - 0.0813B + 128 \end{aligned}$$

2. DCT变换

离散余弦变换（Discrete Cosine Transformation, DCT）具有快速算法，且易于实现等优点，它的快速算法已经可以由专用芯片来实现，因而被广泛采用。

由于图像可以看成是二维数据矩阵，所以图像编码采用二维 DCT 变换。如果将整幅图像看成是一个二维矩阵，则其正交变换的计算量太大，难以实现。所以在实用中，先将一幅图像分割成一个个小图像块。若块选得太小，不利于压缩比的提高。而块越大，计入的相关像素越多，压缩比就越高。但若块过大，不但计算复杂，而且距离较远的像素间相关性减少，压缩比提高不快。所以一般图像典型变换块大小为 8 像素×8 像素或 16 像素×16 像素。

如果原图像为 640×480 的分辨率，划分后的图像将包含 80×60 个这样的子块，如图 5.2 所示。



图 5.2 JPEG 将原图像划分为若干个子块，每个子块包含 8 像素×8 像素

那么 DCT 处理前后数据发生了哪些变化？为讨论问题方便，取图像中一个被放大的 8 像素×8 像素（假设每个像素的灰度值用 8 比特来表示）的子块作为示例，如图 5.3（a）、图 5.3（b）所示。

DCT 对一个 8×8 的图像 $f(x,y)$ （如图 5.3（b）所示）进行处理，得到一个 8×8 的频率系数矩阵 T ，如图 5.3（c）所示。 T 中的每个元素称为变换系数，这些系数都有明确的物理意义， $T(0,0)$ 表示变换域中的低频成分（图像的背景），也称为直流（DC）系数，与图像矩阵（如图 5.3（b）所示）的平均值有关； T 矩阵的其余的值表示变换域中的高频部分（图像的细节），也称为交流（AC）系数。

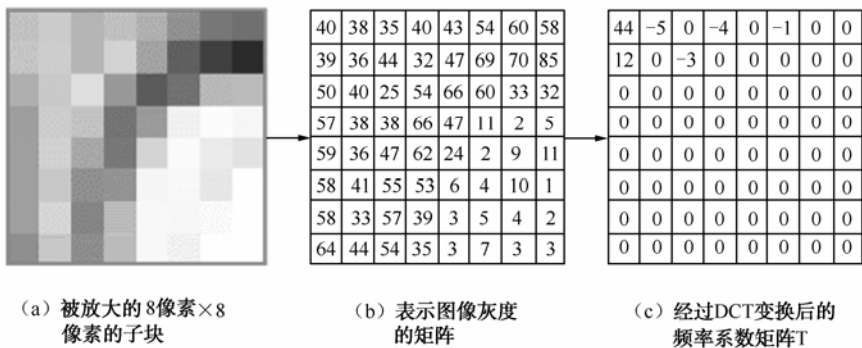


图 5.3 一个 8 像素×8 像素子块的 DCT 变换

从矩阵中可以看出有一个倾向：随着元素离 DC 系数越来越远，它的值就倾向于越来越

小。这意味着通过 DCT 来处理数据，已将图像的代表集结到输出矩阵的左上角的系数中，这个系数就比其他系数携带了更多关于图像的有用信息，同时 DCT 矩阵的右下部分系数几乎不包含有用信息，图像的信息集中在少量系数中，有利于数据压缩。另外，由于大多数图像都由低频信息构成，因此这一点是很有意义的。这样，系数越来越小，而且对描述图像越来越不重要。所以说 DCT 确定了图像的一部分信息，这些信息可被“扔掉”（在矩阵 T 中相应的系数置 0）并且不会对图像的质量带来严重影响，这样能够进一步压缩图像。在图像未进行变换前要实现这一点是难以想象的，当图像在空间域描述时要找出哪些像素对图像的全貌是重要的，而哪些是不重要的是相当困难的。

如果离散余弦变换是不可逆运算（即从 DCT 系数中恢复原始像素信息），那么这种变换是毫无意义的。通过离散余弦变换的逆变换（IDCT）对编码数据进行逆变换，可将图像变回到色度空间，从而实现图像的解压缩，图像得到还原。

3. 量化

DCT 变换的作用是使空间域的能量重新分布，降低图像的相关性。DCT 变换本身并不能达到数据压缩的作用，而要实现图像压缩，就要选择适当的比特分配方案和量化方法。量化的作用是在保证主观图像质量的前提下，丢掉那些对视觉效果影响不大的信息。量化是一种降低精度的过程，所以是有损的，如图 5.4 所示。会产生一定的失真，失真程度视量化等级而定。



图 5.4 经量化后，原图像（左）与 JPEG 压缩图像（右）

8×8 的图像块经过 DCT 变换后，其低频分量都集中在左上角，高频分量分布在右下角（DCT 变换实际上是空间域的低通滤波器）。由于该低频分量包含了图像的主要信息，而高频与之相比，就不那么重要了，所以我们可以忽略高频分量，从而达到压缩的目的。如何将高频分量去掉，这就要用到量化，它是产生信息损失的根源。这里的量化操作，就是将某一个值除以量化表中对应的值。由于量化表左上角的值较小，右上角的值较大，这样就起到了保持低频分量，抑制高频分量的目的。

在 JPEG 标准中采用线性均匀量化器，量化过程为对 $64 (8 \times 8)$ 个 DCT 系数除以量化步长并取整，量化步长由量化表（量化矩阵）决定。量化的计算公式十分简单：

$$\text{量化值}(i, j) = T(i, j) / \text{量化矩阵}(i, j)$$

由公式可见，当量化值较大时，可以保证所有较高频率的分量实际上都将被四舍五入为 0。仅在高频系数很大时才将其编码定为非 0 值，但这种情况很少出现。

在解码过程中，逆量化公式为

$$T(i, j) = \text{量化值}(i, j) \times \text{量化矩阵}(i, j)$$

不难看出，当使用大的量化值时，在逆量化过程中所用的 DCT 输出会有大的误差，幸运的是逆量化过程中高频分量的误差不会对图像的质量有严重影响。显然有许多方案可用来选择量化矩阵中的元素值。

(1) JPEG 推荐的量化表

JPEG 使用的颜色是 YCbCr 模型，Y 分量代表了亮度信息，Cb, Cr 分量代表了色度信息。相比而言，Y 分量更重要一些。对 Y 采用细量化，对 Cb, Cr 采用粗量化，可进一步提高压缩比。所以量化表通常有两张，一张是针对 Y 的，一张是针对 Cr, Cb 的。

JPEG 基本算法包括一套量化表，它是从广泛的实验中得出的。表 5.1 和表 5.2 分别给出了 JPEG 标准所推荐的亮度量化表和色度量化表。量化表中元素为 1~255 的任意整数，其值规定了所对应 DCT 频率系数的量化步长。当频率系数经过量化后，将频率系数由浮点数转变为整数，这才便于执行最后的编码。不过，经过量化阶段后，所有数据只保留整数近似值，也就再度损失了一些数据内容。

【例 5.3】 对如图 5.5 (a) 所示的原图像块进行量化后对图像质量的影响。

计算 DCT 变换之前对原图像中的每个样本数据减去 128，在 IDCT (逆向离散余弦变换) 变换之后对重建图像中的每个样本数据加 128，经过量化的重建图像灰度值 (如图 5.5 (f) 所示) 与原始图像灰度值 (如图 5.5 (a) 所示) 并不完全一致。

表 5.1 亮度量化表

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

表 5.2 色度量化表

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

(2) 量化表的优化

用户可根据人类视觉系统和压缩图像类型的特点对量化表进行优化，并作为编码器的一个输入。选择量化矩阵的好处是在使用 DCT 来压缩图像时控制图像的质量很容易。通过给较大的 DCT 系数选择特别高的量化步长可以获得很好的压缩率，但图像质量很差；通过慎重地选择小的量化步长，压缩率将下降到不是很好的水平，但图像的质量会很好。这样用户有很大的灵活性，从而可根据需要来选择图像的质量。例如，在 Photoshop 中另存文件为 JPG 格式时，用户可对图像的“低、中、高、最佳”品质进行选择，如图 5.6 所示。

139	144	149	153	155	155	155	155	235.6	-1.0	-12.1	-5.2	2.1	-1.7	-2.7	1.3
144	151	153	156	159	156	156	156	-22.6	-17.5	-6.2	-3.2	-2.9	-0.1	0.4	-1.2
150	155	160	163	158	156	156	156	-10.9	-9.3	-1.6	1.5	0.2	-0.9	-0.6	-0.1
159	161	162	160	160	159	159	159	-7.1	-1.9	0.2	1.5	0.9	-0.1	0.0	0.3
159	160	161	162	162	155	155	155	-0.6	-0.8	1.5	1.6	-0.1	-0.7	0.6	1.3
161	161	161	161	160	157	157	157	1.8	-0.2	1.6	-0.3	-0.8	1.5	1.0	-1.0
162	162	161	163	162	157	157	157	-1.3	-0.4	-0.3	-0.5	-0.5	1.7	1.1	-0.8
162	162	161	161	163	158	158	158	-2.6	1.6	-3.8	-1.8	1.9	1.2	-0.6	-0.4

(a) 8×8 的原始图像块

(b) DCT 变换

16	11	10	16	24	40	51	61	15	0	-1	0	0	0	0	0
12	12	14	19	26	58	60	55	-2	-1	0	0	0	0	0	0
14	13	16	24	40	57	69	56	-1	-1	0	0	0	0	0	0
14	17	22	29	51	87	80	62	0	0	0	0	0	0	0	0
18	22	37	56	68	109	103	77	0	0	0	0	0	0	0	0
24	35	55	64	81	104	113	92	0	0	0	0	0	0	0	0
49	64	78	87	103	121	120	101	0	0	0	0	0	0	0	0
72	92	95	98	112	100	103	99	0	0	0	0	0	0	0	0

(c) 使用的量化表

(d) 量化结果

240	0	-10	0	0	0	0	0	144	146	149	152	154	156	156	156
-24	-12	0	0	0	0	0	0	148	150	152	154	156	156	156	156
-14	-13	0	0	0	0	0	0	155	156	157	158	158	157	156	155
0	0	0	0	0	0	0	0	160	161	161	162	161	159	157	155
0	0	0	0	0	0	0	0	163	163	164	163	162	160	158	156
0	0	0	0	0	0	0	0	163	164	164	164	162	160	158	157
0	0	0	0	0	0	0	0	160	161	162	162	162	161	159	158
0	0	0	0	0	0	0	0	158	159	161	161	162	161	159	158

(e) 逆量化结果

(f) IDCT 变换（重建图像）

图 5.5 量化对图像质量的影响



图 5.6 Photoshop 中 JPEG 图像品质的选择

4. 编码

JPEG 压缩算法的最后部分是对量化后的图像进行编码。这一部分由 3 步组成。

(1) 直流系数（DC）编码

因为图像中相邻块之间有很强的相关性，JPEG 标准对 DC 系数采用 DPCM（差分脉冲编码调制，采用预测编码的方式，所谓预测编码就是根据过去的信号样值来预测下一个信号

样值，并仅把预测值与现实样值的差值加以量化，编码）编码方法，即对相邻的 8 像素×8 像素块之间的 DC 系数的差值进行编码。经过 DCT 变换后，低频分量集中在左上角，其中 $T(0, 0)$ （即第一行第一列元素）代表了直流（DC）系数，即 8×8 子块的平均值。由于两个相邻的 8×8 子块的 DC 系数相差很小，所以对它们采用差分编码（DPCM），可以提高压缩比，也就是说，对相邻子块的 DC 系数的差值进行编码。这个差值可用公式表示为

$$\Delta DC_i = DC_i - DC_{i-1}$$

因为 DC 系数常常包含了整幅图像的主要部分，所以这个特殊处理很有意义。

(2) 交流系数（AC）编码

T 矩阵中有 63 个元素是交流（AC）系数，可采用行程编码进行压缩。这 63 个元素采用了“之”字形（Zig-Zag）的排列方法，称为 Z 形扫描。

Z 形扫描算法能够实现高效压缩的原因之一是经过量化后，大量的 DCT 矩阵元素被截成 0，而且零值通常是从左上角开始沿对角线方向分布的，如图 5.5（d）所示。采用行程编码算法（RLE，该编码属于无损压缩编码，基本原理是：一般包含两项，第一项用一个符号串代替具有相同值的连续符号，第二项用来记录原始数据中有多少个这样的值。这样使得编码长度少于自然编码的长度。如 11111000000555555555 的行程编码为：（1，5）（0，6）（5，8），可见，行程编码的位数远远少于原始字符串的位数）沿 Z 形路径可有效地累积图像中的 0 的个数，所以这种编码的压缩效率非常高。Z 形扫描的过程如图 5.7 所示。

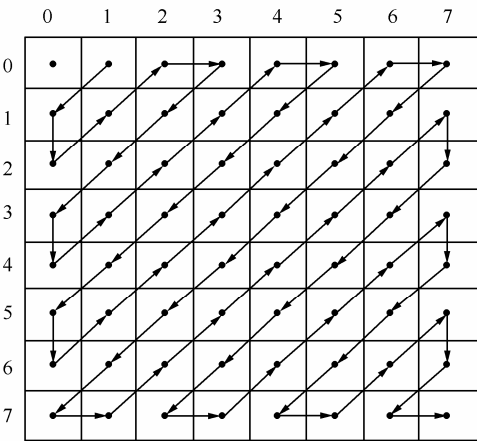


图 5.7 Z 形扫描过程

(3) 熵编码

为了进一步达到压缩数据的目的，需要对 DPCM 编码后的直流系数（DC）和行程编码后的交流系数（AC）再做基于统计特性的熵编码（Entropy Coding，即在编码过程中要求不丢失信息量（熵），是根据消息出现概率的分布特性进行的，属于无损压缩编码）。在 JPEG 有损压缩算法中，使用哈夫曼（Huffman，属于熵编码）编码。哈夫曼编码可以使用很简单的查表（Lookup Table）方法进行编码。在压缩数据符号时，哈夫曼编码对出现频度比较高的符号分配比较短的代码，而对出现频度较低的符号分配比较长的代码。

最后，JPEG 将各种标记代码和编码后的图像数据按帧组成数据流，用于保存、传输和应用。

5.2.3 JPEG2000 简介

JPEG 所具有的优良品质，使它获得极大的成功。然而，随着多媒体应用领域的激增，传统 JPEG 压缩技术也存在着许多不足，无法满足人们对多媒体图像资料的要求。离散余弦变换将图像压缩为 8×8 的小块，然后依次放入文件中，这种算法靠丢弃频率信息实现压缩，因而图像的压缩率越高，频率信息被丢弃得越多。在极端情况下，JPEG 图像只保留了反映图像外貌的基本信息，精细的图像细节都损失了。

JPEG2000 与传统 JPEG 最大的不同，在于它放弃了 JPEG 所采用的以离散余弦变换 (Discrete Cosine Transform) 为主的区块编码方式，区块编码方式的主要缺点是将自然图像中的相关性人为地割裂开来，所以会导致图像还原时出现块与块之间的“边界效应”。而改用以小波转换 (Wavelet Transform) 为主的解析编码方式，小波转换将一幅图像作为一个整体进行变换和编码，很好地保存了图像信息中的相关性，达到了更好的压缩编码效果。

下面说明 JPEG2000 的特点。

1. 高压缩率

JPEG2000 格式的图片压缩比可在 JPEG 基础上再提高 10%~30%，而且压缩后的图像显得更加细腻平滑。如图 5.8 所示是 JPEG2000 和 JPEG 分别采用同样压缩率下的对比效果，可以很明显地看到，JPEG 当被压缩的图像上有大片近似颜色时，会出现马赛克现象，两者的差距是十分明显的。

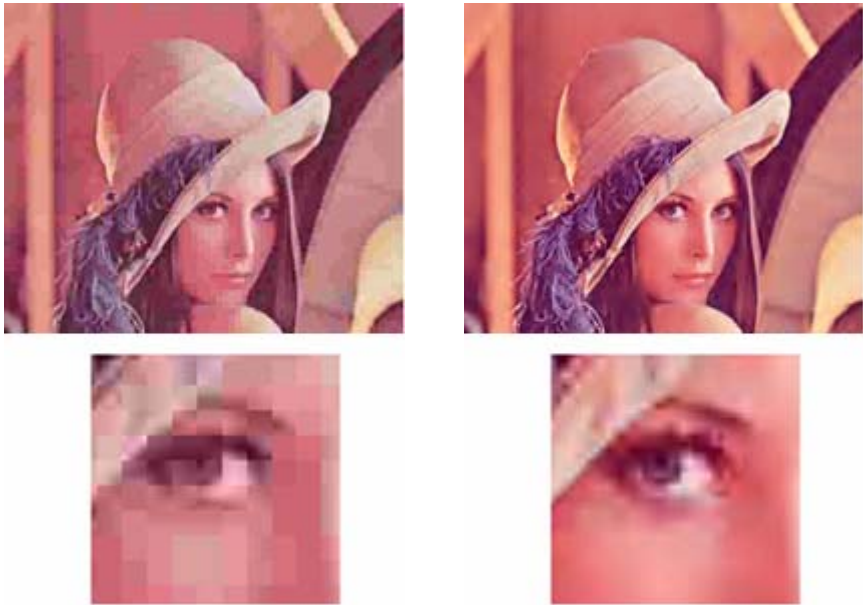


图 5.8 JPEG（左）、JPEG2000（右）图像质量比较

2. 同时支持有损和无损压缩

JPEG 主要支持有损压缩，压缩后数据不能完全复原，而 JPEG2000 则可以同时支持有损和无损两种压缩模式。这为某些要求不能丢失原始信息却又强调较小的文件大小和较高质量

的应用领域（如卫星遥感图像、医学图像、文物照片等）提供了很好的解决方案，因此它更适合保存一些重要图像。

3. 渐进传输

JPEG2000 能实现渐进传输（Progressive Transmission）。它可先传输低分辨率的图像或者图像的轮廓，然后逐步传输其他数据，不断提高图像质量，以满足用户的需要（这一点与 GIF 的“渐显”功能有异曲同工之妙）。这个特性在多媒体网络应用中有重要的意义，例如，当下载一幅图像时，可以比较快地看到这幅图像的概貌，根据对图像质量的要求和可用的网络带宽控制下载图像的数据量。

4. 可对感兴趣区域指定压缩参数

JPEG2000 另一个重要特性是支持感兴趣区域（Region Of Interest, ROI）的编码。用户利用这个特性可指定感兴趣的图像区域，在压缩时对这些图像区域指定特定的压缩质量，或在恢复时选择指定的部分先解压缩，从而使重点突出，这给用户带来了极大的方便。例如，在有些情况下图像中只有一小块区域对用户是有用的，对这些区域采用低压缩比，而其他区域采用高压缩比，在保证不丢失重要信息的同时能有效地压缩数据量，如图 5.9 所示。

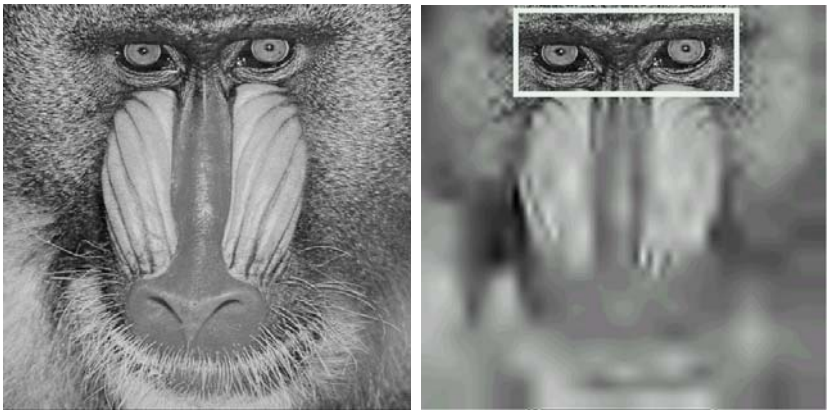


图 5.9 JPEG2000 ROI 编码

5.3 运动图像压缩标准MPEG

MPEG（Moving Picture Experts Group）是运动图像专家组的英文缩写。这个专家组始建于 1988 年，是由 ISO(International Standard Orgnization)与 IEC(International Electronic Committee)于 1988 年联合成立的，专门负责建立视频和音频数据压缩标准。JPEG 和 MPEG 都是在 ISO 领导下的专家小组，其成员也有很大的交叠。JPEG 的目标集中于静止图像压缩，而 MPEG 的目标是针对活动图像的数据压缩，但静止图像与活动图像有密切关系。

MPEG 标准主要有 MPEG—1，MPEG—2，MPEG—4，MPEG—7 和 MPEG—21。

5.3.1 MPEG—1

MPEG—1 处理的是标准图像交换格式（Standard Interchange Format，SIF）或者称为源输入格式（Source Input Format，SIF）的电视，即 NTSC 制为 352×240，30fps，PAL 制为 352×288，25fps，压缩的输出速率定义在 1.5 Mbps 以下，提供了相当于家用录像系统（VHS）的影音质量。这个标准主要是针对当时具有这种数据传输率的 CD-ROM 和网络而开发的，用于在 CD-ROM 上存储数字影视和在网络上传输数字影视。另外，一些数码相机和数字摄像机把 MPEG—1 作为其数字音视频的记录格式，以 MPEG—1 作为音视频编码标准的 VCD 在我国已比较普及。

MPEG—1 标准体系共分为 5 个部分，其中，系统、视频和声音 3 部分，1993 年获得通过成为标准；第 4 部分一致性测试 1995 年通过；第 5 部分软件仿真为技术报告。

第 1 部分（系统，Systems）：规定视频数据、声音数据及其他相关数据的同步。

第 2 部分（视频，Video）：规定视频数据的编码和解码。

第 3 部分（音频，Audio）：规定声音数据的编码和解码。

第 4 部分（一致性测试，Conformance Testing）：这个标准详细说明如何测试比特数据流（Bitstreams）和解码器是否满足 MPEG—1 前 3 个部分中所规定的要求。编码器制造商和客户均可使用这些方法来验证编码器产生的码流是否正确。

第 5 部分（软件仿真，Software Simulation）：从技术上讲，这部分不算标准，只是一种技术报告，描述了 MPEG—1 标准的前 3 部分功能的软件实现，源代码是不公开的。

MPEG—1 编码的实施框图如图 5.10 所示。

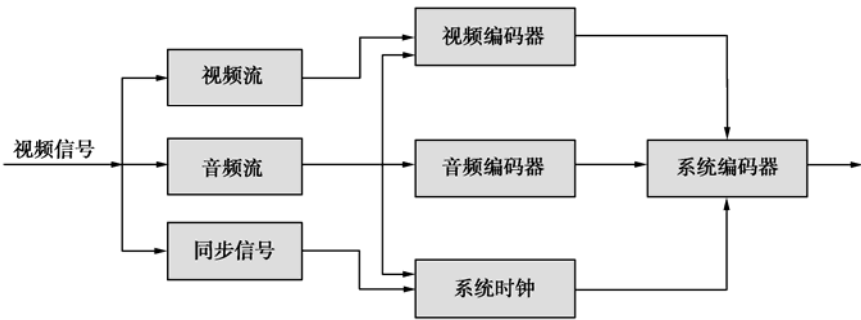


图 5.10 MPEG—1 编码的实施框图

1. 音频编码

MPEG 声音（ISO/IEC 11172—3）压缩算法是世界上第一个高保真声音数据压缩国际标准，并且得到了极其广泛的应用。虽然 MPEG 声音标准是 MPEG 标准的一部分，但它也完全可以独立应用。

MPEG 编码器的输入信号为线性 PCM 信号，采样率为 32kHz，44.1kHz 或 48kHz，输出为 32~384kbps，如图 5.11 所示。



图 5.11 MPEG 音频编码器的输入/输出

MPEG 声音标准提供了 3 个独立的压缩层次: 层 1(Layer 1)、层 2(Layer 2)和层 3(Layer 3), 用户对层次的选择可在复杂性和声音质量之间进行权衡, 在尽可能保持 CD 音质为前提的条件下, MPEG 声音标准一般所能达到的压缩率见表 5.3。层 1 (Layer 1) 的编码较简单, 编码器的输出数据率为 384kbps, 主要用于数字盒式录音磁带; 层 2 (Layer2) 的算法复杂度中等, 编码器的输出数据率为 256~192 kbps, 其应用包括数字音频广播 (DAB) 和 VCD 等; 层 3 (Layer3) 的编码较复杂, 编码器的输出数据率为 128~112kbps, 主要用于因特网上高质量声音的传输。最近几年流行起来的所谓“MP3 音乐”就是一种采用 MPEG—1 层 3 编码的高质量数字音乐, 它能以 10 倍左右的压缩比降低高保真数字音乐的存储量, 使一张普通的 CD 光盘上可以存储大约 100 首 MP3 歌曲。

表 5.3 MPEG—1 声音压缩标准

名 称	压缩后的码率	压缩率	算法复杂度	声道数目	主要应用
MPEG—1 层 1	384kbps	4 : 1	低	2	数字盒式录音带
MPEG—1 层 2	256~192kbps	6 : 1~8 : 1	中	2	DAB, VCD
MPEG—1 层 3	128~112kbps	10 : 1~12 : 1	高	2	Internet、MP3 音乐

2. 视频编码

MPEG—1 压缩的核心是视频, 为满足高压缩比和随机访问两方面要求, MPEG—1 标准采用了一系列技术以获得高压缩比。

MPEG—1 视频压缩技术基本方法可以归纳成两个要点: ① 在空间方向上, 图像数据压缩采用 JPEG (Joint Photographic Experts Group) 压缩算法来去掉冗余信息。② 在时间方向上, 图像数据压缩采用运动补偿 (Motion Compensation) 算法来去掉冗余信息。

1) MPEG—1 视频图像的类型

为了保证图像质量基本不降低而又能够获得高的压缩比,MPEG 专家组定义了 3 种图像: 帧内图像 I (Intra), 预测图像 P (Predicted) 和双向预测图像 B (Bidirectionally Interpolated), 典型的排列如图 5.12 所示。这 3 种图像采用 3 种不同的算法进行压缩。

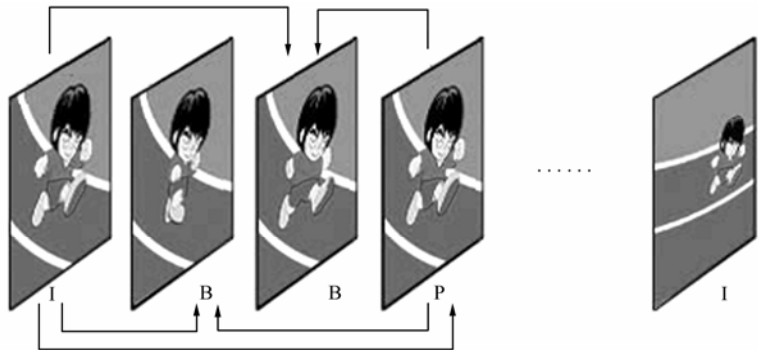


图 5.12 MPEG 专家组定义的 3 种图像的典型排列

(1) 帧内图像 I 的压缩编码算法

帧内图像 I 是一个独立的帧 (相当于关键帧), 其信息由自身画面决定, 不需要参照其他

画面产生，是预测图像 P 和双向预测图像 B 的参考图，压缩编码采用类似 JPEG 压缩算法。

(2) 预测图像 P 的压缩编码算法

预测图像 P，它参照前一幅 I 图像或 P 图像产生，编码以图像宏块（Macroblock）为基本编码单元，一个宏块定义为 I 像素 $\times J$ 像素的图像块，一般取 16×16 。预测图像 P 使用两种类型的参数来表示：一种参数是当前要编码的图像宏块与参考图像的宏块之间的差值；另一种参数是宏块的移动矢量。移动矢量的概念如图 5.13 所示。

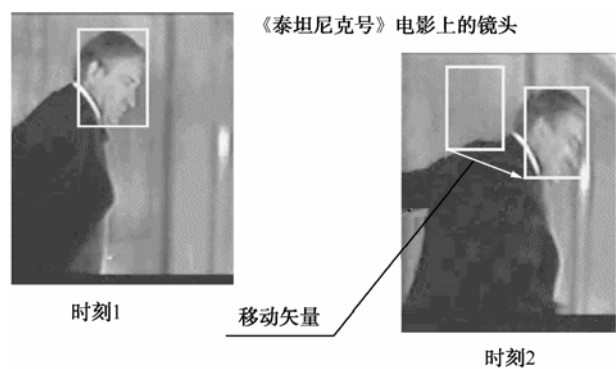


图 5.13 移动矢量的概念

求解差值的方法如图 5.14 (a) 所示。假设编码图像宏块 M_{PI} 是参考图像宏块 M_{RJ} 的最佳匹配块，它们的差值就是这两个宏块中相应像素值之差。对所求得的差值进行彩色空间转换，并做 4:1:1 的子采样得到 Y, Cb 和 Cr 分量值，然后仿照 JPEG 压缩算法对差值进行编码，计算出的移动矢量也要进行哈夫曼编码。

求解移动矢量的方法定义在图 5.14 (b) 中。在求两个宏块差值之前，需要找出编码图像中的预测图像编码宏块 M_{PI} 相对于参考图像中的参考宏块 M_{RJ} 所移动的距离和方向，这就是移动矢量（Motion Vector）。

要使预测图像的精度更高，就要求找到与参考宏块 M_{RJ} 最佳匹配的预测图像编码宏块 M_{PI} 。所谓最佳匹配是指这两个宏块之间的差值最小。

从以上分析可知，对预测图像的编码实际上就是寻找最佳匹配图像宏块，在整个 MPEG 视频图像压缩过程中，寻找最佳匹配宏块要占据相当多的计算时间，匹配得越好，重构的图像质量越高。为减少搜索次数，现在已开发出许多简化算法用来寻找最佳宏块。

(3) 双向预测图像 B 的压缩编码算法

双向预测图像 B，它参照前一幅和后一幅 I 图像或 P 图像产生，压缩编码框图如图 5.15 所示。具体计算方法与预测图像 P 的算法类似，这里不再重复。

2) MPEG—1 视频图像的结构

MPEG 编码器算法允许选择帧内图像 I 的频率和位置。帧内图像 I 的频率是指每秒钟出现帧内图像 I 的次数，位置是指时间方向上帧所在的位置。一般情况下，帧内图像 I 的频率为 2。MPEG 编码器也允许在一对帧内图像 I 或者预测图像 P 之间选择双向预测图像 B 的数目。帧内图像 I、预测图像 P 和双向预测图像 B 数目的选择依据主要是根据节目的内容。例如，对于快速运动的图像，帧内图像 I 的频率可以选择高一些，而双向预测图像 B 的数目可以选择少一点；对于慢速运动的图像帧内图像 I 的频率可以低一点，而双向预测图像 B 的数目可以选择多一点。此外，在实际应用中还要考虑媒体的速率。

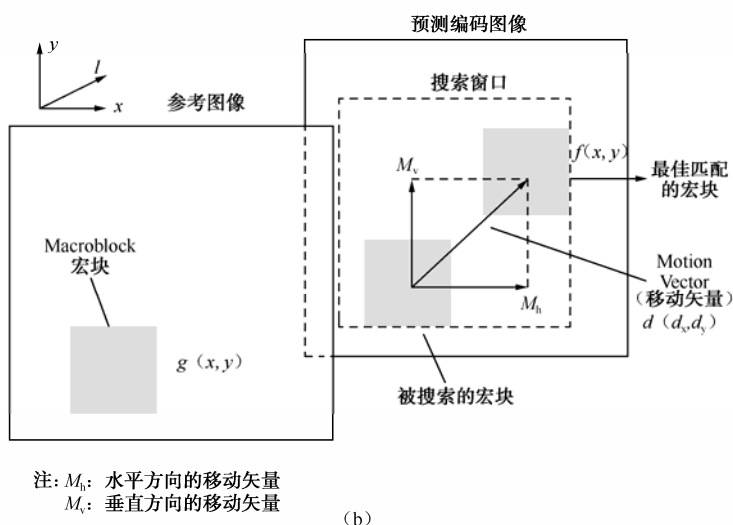
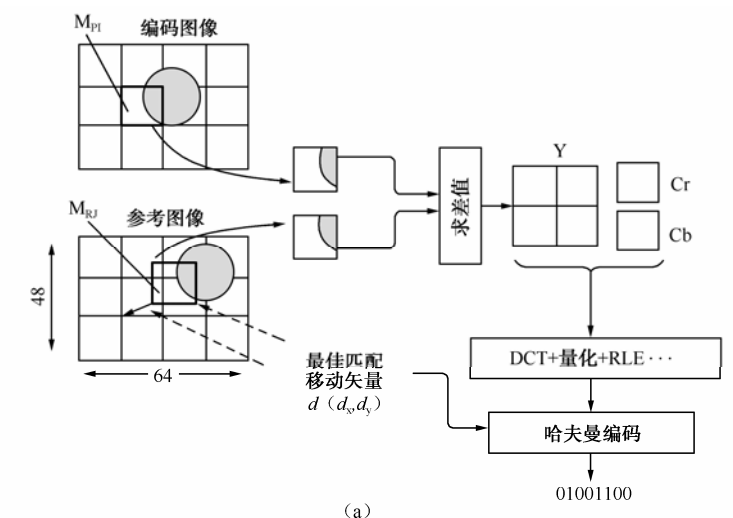


图 5.14 预测图像 P 的压缩算法框图

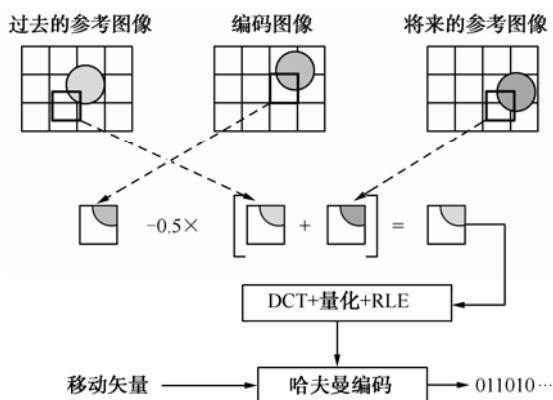


图 5.15 双向预测图像 B 的压缩编码算法框图

一个典型的帧内图像 I、预测图像 P 和双向预测图像 B 的安排如图 5.16 所示。编码参数为：帧内图像 I 的距离为 $N=15$ ，预测图像 P 的距离为 $M=3$ 。

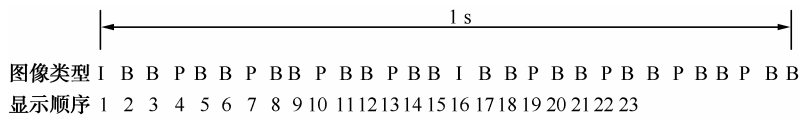


图 5.16 MPEG 视频图像帧编排

帧内图像 I 采用帧内编码方式，即只利用了单帧图像内的空间相关性，而没有利用时间相关性。由于帧内图像 I 不依赖其他帧，所以是随机存取的入点，同时也是解码的基准帧。帧内图像 I 主要用于视频的初始化和获取，以及节目的切换和插入，帧内图像 I 的压缩倍数相对较低。

预测图像 P 和双向预测图像 B 图像采用帧间编码方式，即同时利用了空间和时间上的相关性。预测图像 P 只采用前向时间预测，可以提高压缩效率和图像质量。预测图像 P 中可以包含帧内编码的部分，即预测图像 P 中的每一个宏块可以是前向预测，也可以是帧内编码。双向预测图像 B 图像采用双向时间预测，可以大大提高压缩倍数。值得注意的是，由于双向预测图像 B 图像采用了未来帧作为参考，因此 MPEG—1 编码码流中图像帧的传输顺序和显示顺序是不同的。图 5.16 显示的是播放时看到的图像次序，而不是传送的图像次序。预测图像 P 在最初的两个双向预测图像 B 前面传送，而第二个帧内图像 I 在最后的两个双向预测图像 B 前面传送。然后预测图像 P 和两个帧内图像 I 可以被缓存起来，这样接下来收到的双向预测图像 B 就可以在观看端进行解码。

从压缩的程度来看，帧内图像 I 的压缩率最小；由于预测图像 P 只存储当前帧和参考帧的误差信号，因此预测图像 P 得到了较大的压缩；而双向预测图像 B 的压缩率是最大的，增加双向预测图像 B 的数目能提高压缩比，但视频质量会有损失。

5.3.2 MPEG—2

MPEG—2 标准从 1990 年开始研究，1994 年由 ISO/IEC 制定公布，它是一个高质量视频和声音编码标准。MPEG—2 可以说是 MPEG—1 的扩充，因为它们的基本编码算法都相同。但 MPEG—2 不是 MPEG—1 的简单升级，MPEG—2 在系统和传输方面做了更加详细的规定和进一步的完善。MPEG—2 要达到的最基本目标是：位速率为 4~9 Mbps，最高达 15 Mbps。

MPEG—2 标准分为 9 个部分，各部分的内容描述如下。

第 1 部分（系统，Systems）：规定视频数据、声音数据及其他相关数据的同步。

第 2 部分（视频，Video）：规定视频数据的编码和解码。

第 3 部分（音频，Audio）：规定声音数据的编码和解码，是 MPEG—1 Audio 的扩充，支持多个声道，向下兼容 MPEG—1 Audio 标准。

第 4 部分（一致性测试，Conformance Testing）：描述测试一个编码码流是否符合 MPEG—2 码流的方法。

第 5 部分（软件仿真，Software Simulation）：描述了 MPEG—2 标准的第 1，2，3 部分的软件实现方法。

第 6 部分（数字存储媒体命令和控制扩展协议，DSM—CC，Digital Storage Media Command and Control）：用于管理 MPEG—1 和 MPEG—2 的数据流，使数据流既可在单机上运行，又可在异构网络环境下运行。

以上 6 个部分均已获得通过，成为正式的国际标准，并在数字电视等领域中得到了广泛的实际应用。此外，MPEG—2 标准还有 3 个部分：第 7 部分规定与 MPEG—1 音频不兼容的多通道音频编码（Advanced Audio Coding，AAC）；第 8 部分原计划用于采样精度为 10 位的视频编码，但由于工业界对此兴趣不大，现已停止；第 9 部分规定了传送码流的实时接口。

1. MPEG—2 的音频编码

MPEG—2 标准委员会定义了两种声音数据压缩格式，一种称为 MPEG—2 Audio，或者称为 MPEG—2 多通道（Multichannel）声音，因为它与 MPEG—1 Audio 是兼容的，所以又称为 MPEG—2 BC（Backward Compatible）。另一种称为 MPEG—2 AAC（Advanced Audio Coding），因为它与 MPEG—1 声音格式不兼容，因此通常称为非向下兼容 MPEG—2 NBC（Non-Backward-Compatible）标准。

(1) MPEG—2 Audio

MPEG—2 Audio（ISO/IEC 13818—3）和 MPEG—1 Audio（ISO/IEC 1117—3）标准都使用相同种类的编译码器，层 1、层 2 和层 3 的结构也相同。MPEG—2 声音标准与 MPEG—1 标准相比，MPEG—2 做了如下扩充：①增加了 16kHz，22.05kHz 和 24kHz 采样频率；②扩展了编码器的输出速率范围，由 32~384kbps 扩展到 8~640kbps；③增加了声道数，支持 5.1 声道和 7.1 声道的环绕声。此外 MPEG—2 还支持 Linear PCM(线性 PCM)和 Dolby AC3(Audio Code Number 3) 编码。它们的差别见表 5.4。

表 5.4 MPEG—1 和 MPEG—2 的声音数据规格差别

参数名称	采样频率（kHz）	每个样本的比特数	最大数据传输率（bps）	最大声道数
Linear PCM	48/96	16/20/24	6.144M	8
Dolby AC3	32/44.1/48	压缩（16bit）	448k	5.1
MPEG—1 Audio	32/44.1/48	压缩（16bit）	32~448k	2
MPEG—2 Audio	16/22.05/24/32/44.1/48	16	8~640k	5.1/7.1

MPEG—2 Audio 的“5.1 环绕声”也称为“3/2-立体声加 LFE”，其中的“.1”就是指 LFE 声道。它的含义是播音现场的前面可有 3 个喇叭声道（左、中、右），后面可有 2 个环绕声喇叭声道，LFE（Low Frequency Effects）是低频（3~120Hz）音效的加强声道，如图 5.17（a）所示。7.1 声道环绕立体声与 5.1 类似，如图 5.17（b）所示。

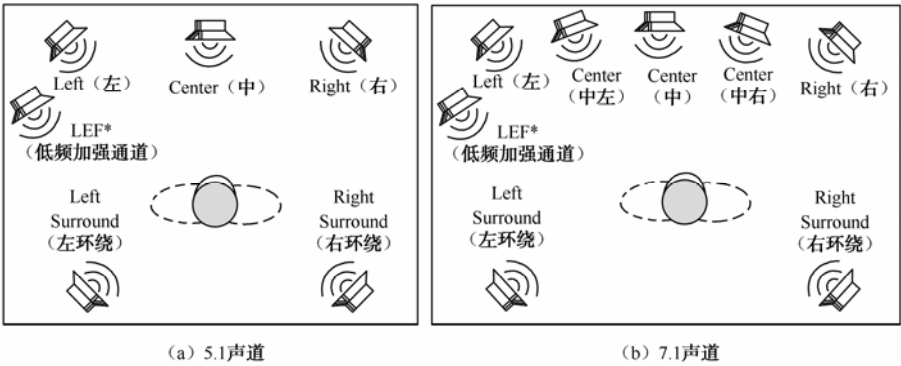


图 5.17 立体环绕声

Dolby AC3 支持 5.1 声道（左、中、右、左环绕、右环绕和 0.1kHz 以下的低音音效声道），声音样本的精度为 20 位，每个声道的采样频率可以是 32kHz，44.1kHz 或者 48kHz。

MPEG—2 标准的第 3 部分是 MPEG—1 声音标准的扩展，扩展部分就是多声道扩展（Multichannel Extension），这个标准称为 MPEG—2 向下兼容多声道声音编码（MPEG—2 Backwards Compatible Multichannel Audio Coding）标准，简称为 MPEG—2 BC。

(2) MPEG—2 AAC

MPEG—2 AAC 是 MPEG—2 标准中的一种非常灵活的声音感知编码标准。就像所有感知编码一样，MPEG—2 AAC 主要使用听觉系统的掩蔽特性来减少声音的数据量，并且通过把量化噪声分散到各个子带（频段）中，用全局信号把噪声掩蔽掉。

AAC 支持的采样频率可从 8kHz 到 96kHz，AAC 编码器的音源可以是单声道的、立体声的和多声道的声音。AAC 标准可支持 48 个主声道、16 个低频音效加强通道 LFE（Low Frequency Effects）、16 个配音声道（Overdub Channel）或者叫做多语言声道（Multilingual Channel）和 16 个数据流。MPEG—2 AAC 在压缩比为 11：1，即每个声道的数据率为 $(44.1 \times 16) / 11 = 64\text{kbps}$ ，而 5 个声道的总数据率为 320 kbps 的情况下，很难区分还原后的声音与原始声音之间的差别。与 MPEG—1 的层 2 相比，MPEG—2 AAC 的压缩率可提高 1 倍，而且质量更高，与 MPEG—1 的层 3 相比，在质量相同的条件下数据率是它的 70%。

AT&T 公司对 MPEG—2 AAC 技术进行了一些改良，增加了最关键的音乐传播认证技术，这样就形成了 MP4 音乐。虽然 MP4（压缩比为 15：1~20：1）有很多先进的优点，但在与 MP3（压缩比为 12：1）竞争的过程中鹿死谁手还有待分晓。关键原因是 MP3 的使用没有任何限制，这种特性在本来就没有限制的网络上得到充分体现。相反的是虽然 AAC 技术是公开的，但 MP4 本身却是受到严格的专利和使用许可证保护的，毫无疑问这肯定使得 MP4 的应用范围非常狭窄。

2. MPEG—2 的视频编码

在 MPEG—2 标准化阶段，考虑到要适应不同数据速率设备的应用，MPEG 专家组定义了 3 种质量不同的编码方式：信噪比可变性（Signal-to-Noise Scalability）、空间分辨率可变性（Spatial Scalability）和时间分辨率可变性（Temporal Scalability）。

信噪比可变性是指图像质量的折中，对于数据率比较低的解码器使用比较低的信噪比，而对数据率比较高的解码器则使用比较高的信噪比；空间分辨率可变性是指图像的空间分辨率的折中，对于低速率的接受器使用比较低的图像分辨率，而对于数据率比较高的接受器使用比较高的图像分辨率；时间分辨率可变性是指图像在时间方向上分辨率的折中，与空间分辨率类似。

MPEG—2 为此引入了“配置（Profiles）”和参数“等级（Levels）”的概念。每种配置定义一套新的算法，而每一个等级指定一套参数范围（如图像大小、帧速率和位速率）。MPEG—2 的配置规格见表 5.5，表 5.6 列出了该等级下允许的最高分辨率及帧频率上限。

表 5.5 MPEG—2 的配置规格

配置（Profile）	算法（Algorithms）
High（HP）	支持由空间分辨率可变配置（Spatial Scalable Profile）提供的所有功能和其他规定功能 子采样格式：YUV，4：2：2 用于进一步提高图像质量

配置（Profile）	算法（Algorithms）
Spatial Scalable（SSP）	支持信噪比可变配置（SNR Scalable Profile）提供的所有功能和空间分辨率可变（Spatial scalable coding）算法（2 层） 子采样格式：YUV，4：2：0
SNR Scalable（SNRP）	支持基本配置（Main Profile）提供的所有功能和信噪比可变编码（SNR scalable coding）算法（2 层） 子采样格式：YUV，4：2：0
Main（MP）	非可变速率编码算法支持随机存取，B 图像预测方式 子采样格式：YUV，4：2：0
Simple（SP）	除不支持基本配置（Main Profile）提供的 B 图像预测功能外，基本配置的其他所有功能都支持 子采样格式：YUV，4：2：0

表 5.6 MPEG—2 的等级

等级（Level）	参数（Parameters）	说明
High （HL）	1920samples/line	1920 样本/行
	1152lines/frame	1152 行/帧
	60fps	60 帧/秒
	80Mbps	80 兆比特/秒
High 1440 （H14L）	1440samples/line	1440 样本/行
	1152lines/frame	1152 行/帧
	60fps	60 帧/秒
	60Mbps	60 兆比特/秒
Main （ML）	720samples/line	720 样本/行
	576lines/frame	576 行/帧
	30fps	30 帧/秒
	15Mbps	15 兆比特/秒
Low （LL）	352samples/line	352 样本/行
	288lines/frame	288 行/帧
	30fps	30 帧/秒
	4Mbps	4 兆比特/秒

由配置（Profile）和参数等级（Level）组合起来的 MPEG—2 所支持的各种电视规格见表 5.7（并不是所有的 Profile/Level 组合都是有意义的）。前者定义质量的可变性（Scalability）和彩色空间分辨率的句法子集，后者定义图像分辨率和每种配置的最大位速率的参数集。例如，当前使用得最普遍的描述符是 MP@ML（Main Profile，Main Level）（标准数字电视和 DVD 视频采用），可译成“基本配置@基本级电视”或者“基本句子子集@基本参数级”，它指的是具有这种特性的电视：帧速率为 30fps，分辨率为 720×576×30，子采样格式为 4：2：0，位速率达 15Mbps。MPEG—2 标准期待大多数 MPEG—2 设备都能够支持这种电视。又如，

MP@HL（Main Profile, High Level）描述符指的是帧速率为 30fps、分辨率为 1920×1152×60、子采样格式为 4：2：0、位速率达 80Mbps 的 HDTV（高清晰电视）。

表 5.7 MPEG—2 配置等级和参数级

Level/Profile (等级/配置)	Simple (SP)	Main (MP)	SRN Scalability (SNRP)	Spatial Scalability (SSP)	High (HP)
High (HL)		4：2：0 1920×1152×60 60Mbps I, P, B			4：2：0, 4：2：2 1920×1152×60 80Mbps I, P, B
High-1440 (H14L)		4：2：0 1440×1152×60 60Mbps I, P, B		4：2：0 1440×1152×60 60Mbps I, P, B	4：2：0, 4：2：2 1440×1152×60 60Mbps I, P, B
Main (ML)	4：2：0 720×576×30 15Mbps I, P	4：2：0 720×576×30 15Mbps I, P, B	4：2：0 720×576×30 15Mbps I, P, B		4：2：0 720×576×30 20Mbps I, P, B
Low (LL)		4：2：0 352×288×30 4Mbps I, P, B	4：2：0 352×288×30 4Mbps I, P, B		

5.3.3 MPEG—4

1. MPEG—4 概述

MPEG—4 是 ISO 公布的“超低比特率活动图像和语音压缩标准”，1999 年 MPEG—4 正式成为国际标准，并在 2000 年推出了 MPEG—4 Version 2.0。MPEG—4 是针对低比特率下的视频、音频编码的，更加注重多媒体系统的交互性和灵活性。为此，MPEG—4 引入了 AV 对象（Audio/Visual Objects），使得更多的交互操作成为可能。

MPEG—4 不再将图像看成是一个矩形像素阵列的序列，不再将音频看成是一个多声道或单声道的声音，而是深入到组成一个场景的视频、音频对象的语义中去，对不同的主体采用不同的编码方式。例如，把一幅图像中活泼的白猫和毛线团及背景中的房间分别作为 AV 对象进行编码。因此 MPEG—4 是完全基于对象的一种编码方式，它具有高效编码、高效存储与传播，以及可交互操作的特性。

MPEG—4 对 AV 对象的操作主要有：采用 AV 对象来表示听觉、视觉或者视听组合内容；组合已有的 AV 对象来生成复合的 AV 对象，并由此生成 AV 场景；对 AV 对象的数据灵活地进行多路合成与同步，以便选择合适的网络来传输这些 AV 对象数据；允许接收端的用户在 AV 场景中对 AV 对象进行交互操作等。如图 5.18 所示描述了一个 MPEG—4 视频终端根据对象及场景描述重建一个场景的例子。

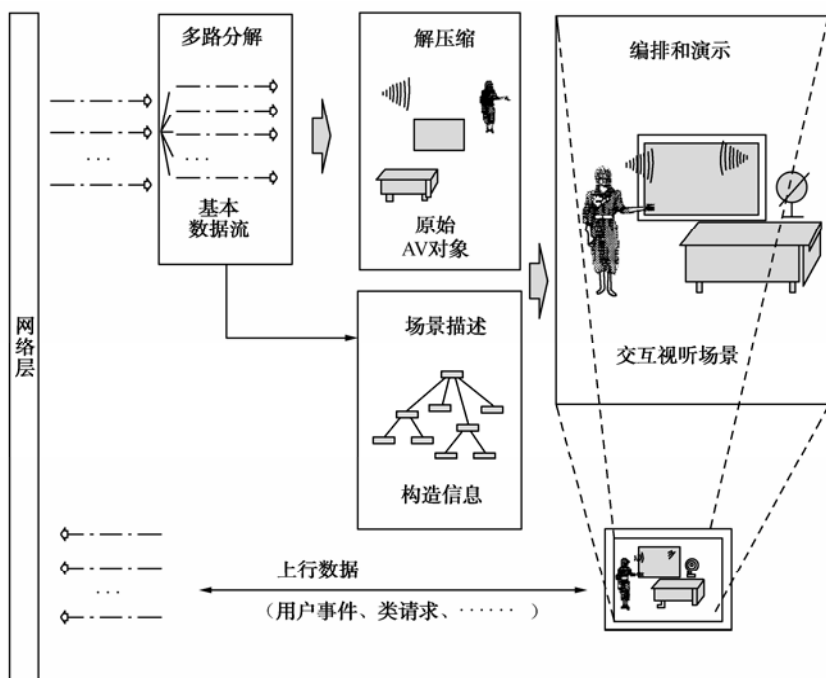


图 5.18 MPEG—4 接收端模型

MPEG—4 标准由以下 6 个主要部分构成。

第 1 部分（系统，Systems）：规定视频数据、声音数据及其他相关数据的同步。

第 2 部分（视频，Video）：规定视频数据的编码和解码。MPEG—4 支持对自然和合成的视觉对象的编码。合成的视觉对象包括 2D，3D 动画和人面部表情动画等。

第 3 部分（音频，Audio）：规定声音数据的编码和解码。与视频编码类似，MPEG—4 不仅支持自然声音，而且支持合成声音。

第 4 部分（一致性测试，Conformance testing）：描述测试一个编码码流是否符合 MPEG—4 码流的方法。

第 5 部分（参考软件，Reference Software）：提供了用于演示功能和说明本标准其他部分功能的软件。

第 6 部分（多媒体传输整体框架，The Delivery Multimedia Integration Framework, DMIF）：它用来管理多媒体数据流。通过 DMIF，MPEG—4 可以建立起具有特殊服务质量（QoS）的信道和面向每个基本流的带宽。DMIF 覆盖了 3 种主要技术：广播技术、交互网络技术和光盘技术，如图 5.19 所示。

与 MPEG—1 和 MPEG—2 相比，MPEG—4 更适于交互 AV 服务及远程监控，它的设计目标使其具有更广的适应性和可扩展性，可以利用很窄的带宽通过帧重建技术压缩和传输数据，从而能以最少的数据获得最佳的图像质量。因此，它将在数字电视、动态图像、互联网、实时多媒体监控、移动多媒体通信、Internet/Intranet 上的视频流与可视游戏、DVD 上的交互多媒体应用等方面大显身手。

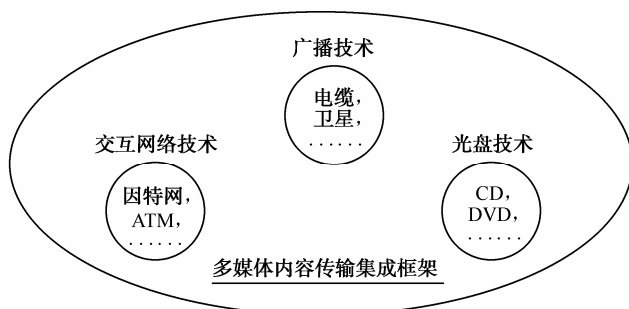


图 5.19 DMIF 覆盖的 3 种主要技术

2. MPEG—4 音频编码

MPEG—4 Audio 标准可集成从语音到高质量的多通道声音，从自然声音到合成声音，编码方法还包括参数编码 (Parametric Coding)，码激励线性预测 (Code Excited Linear Predictive, CELP) 编码、时间/频率 T/F (Time/Frequency) 编码、结构化声音 SA (Structured Audio) 编码和文本-语音 TTS (Text-To-Speech) 系统的合成声音等。

3. MPEG—4 的视频编码

MPEG—4 Video 编码算法支持由 MPEG—1 和 MPEG—2 提供的所有功能，包括对各种输入格式下的标准矩形图像、帧速率、位速率和隔行扫描图像源的支持。MPEG—4 Video 算法的核心是支持基于内容 (Content-Based) 的编码和解码功能，也就是对场景中使用分割算法抽取的单独的物理对象进行编码和解码。

MPEG—4 采用基于内容编码方法的一个重要优点是，使用合适的和专门的基于对象移动预测工具 (Object-based Motion Prediction Tools) 可以明显提高场景中某些电视图像对象的压缩效率。

如图 5.20 所示为 MPEG—4 对视频图像序列进行编码的一个实际例子。左上角的图是背景全景图。右上角的图是一个没有背景的子图像全景图，可以把网球运动员当做一个视频对象 (VO)，一般把这种可以独立移动的小图像称为子图像 (Sprite)。下面的图是接收端合成的全景图。在编码之前这个子图像全景图从背景全背景图序列中抽出来，然后分别对它们进行编码、传送和解码，最后再合成。

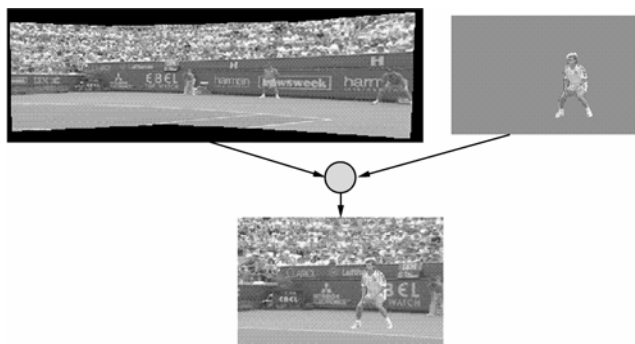


图 5.20 MPEG—4 视频序列编码实例

4. MPEG—4 编码器

1998 年,微软开发了第一个在 PC 上使用的 MPEG—4 编码器,它包括 MS MPEG—4 V1, MS MPEG—4 V2, MS MPEG—4 V3 的系列编码内核。前面两种都可以用来制作 AVI 文件,至今都作为 Windows 的默认组件。不过 V1 和 V2 的编码质量都还不太好,直到 MS MPEG—4 V3 开始,画面质量有了显著的进步。不过微软却决定仅将这个 MS MPEG—4 V3 的视频编码内核封闭在 Windows Media 流媒体技术中,也就是我们熟知的 ASF 文件之中,不能再用于 AVI 文件。ASF 文件虽然有一些好处,但是过于封闭甚至不能被编辑。因此,很快便有人修改了微软的 MS MPEG—4 V3,解除了不能用于 AVI 文件的限制,并开放了其中一些压缩参数,由此,也就诞生了今天所熟悉的 MPEG—4 编解码器 DivX;-)3.11,它是目前互联网上普遍采用的 MPEG—4 编码器之一。为了强化 MPEG—4 核心的开放,最近网上又出现了一种新的 MPEG—4 编码器——XviD,被认为是目前最快的 MPEG—4 编码器。

5. MPEG—4 的应用

首先介绍网络通信和个人应用方面。MPEG—4 的压缩范围是可调整的。MPEG—1 和 MPEG—2 都是对某一个比特率和应用进行优化。MPEG—1 是在 1.5Mbps 左右,MPEG—2 的 DVD 是在 2Mbps~10Mbps,数字 TV 是在 17~20Mbps 左右。MPEG—4 的范围很大,最低几乎可以到 0,最高值可达 50Mbps。这主要是由于 MPEG—4 采用了基于对象的编码和分级编码等关键技术。这样 MPEG—4 在视频通信方面是很有用的,不管是局域网、无线网,还是广域网等,都可以用 MPEG—4。另外,它是独立于任何设备的,所以 MPEG—4 可以用在笔记本电脑、掌上计算机或机顶盒上。

第二,由于 MPEG—4 是面向对象的,所以可以用于视频会议。例如,10 个人在全球的不同地点召开视频会议,MPEG—4 可以把每个人的形象取出来,安排在一起,就像 10 个人同在一张桌子上开会。

第三,MPEG—4 的可伸缩性很好,所以对 Internet 的传输是非常理想的。我们知道 Internet 的带宽是不固定的,可以变来变去,而且经常丢失资料包。MPEG—4 具有的分级编码和错误恢复功能可以提高 Internet 的鲁棒性。

第四,MPEG—4 是基于对象的,它不仅可以压缩自然视频,而且可以压缩人工合成的图形或视频,这可以用于节目的制作。这样电影制片厂和电视台都可以用 MPEG—4 集成各种各样的媒体信息。

5.3.4 MPEG—7

继 MPEG—4 之后,要解决的矛盾就是对日渐庞大的图像、声音、视频信息的管理和检索。针对这个矛盾,MPEG 提出了解决方案——MPEG—7。MPEG—7 于 1996 年 10 月开始研究,确切地讲,MPEG—7 并不是一种压缩编码方法,其正规的名字叫做“多媒体内容描述接口(Multimedia Content Description Interface)”,其目的是制定一套描述符标准,用来描述各种类型的多媒体信息及它们之间的关系,以便更快更有效地检索信息,如图 5.21 所示。

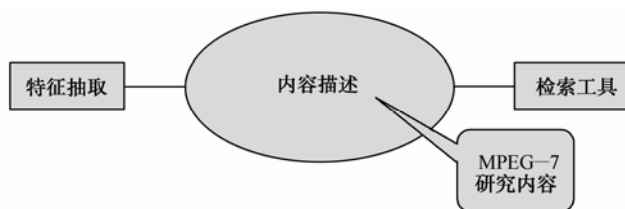


图 5.21 MPEG—7 的研究内容

从图 5.21 中可以看出，MPEG—7 是建立在特征抽取之上的。特征的自动分析和抽取对 MPEG—7 是至关重要的，抽象程度越高，自动抽取就越困难，而且不是都能够自动抽取的，因此开发自动的和交互式半自动抽取的算法和工具都是很有用的。尽管如此，特征抽取和检索工具都不包含在 MPEG—7 标准中，这样使得 MPEG—7 具有很强的通用性和可扩展性，不仅充分利用现有的技术，还给未来的发展提供了足够的空间。

1. MPEG—7 标准系统结构

MPEG—7 标准主要由以下几个主要部分构成。

第 1 部分（系统，Systems）：包括对描述符进行高效传输和存取的工具、对内容和描述进行同步的工具，以及管理和保护知识产权的工具。

第 2 部分（描述定义语言，Description Definition Language，DDL）：定义 MPEG—7 描述工具的语法，定义新的描述方案。

第 3 部分（视频，Video）：包括用来描述视频对象的各种描述符和描述方案。它的组成部分包括其基本的结构和涉及几种基本的视觉特征描述符——颜色、纹理、形状、运动、定位及其他，每一类都有基本和复杂的描述符和描述方案组成。

第 4 部分（音频，Audio）：处理音频描述。包括音频描述结构、声音效果描述工具、乐器音质描述工具、语音识别描述工具、带有旋律的描述工具等。

第 5 部分（多媒体描述方案，Multimedia Description Schemes，MMDS）：用于规范多媒体描述符的生成和不同描述符之间的有机联系。

第 6 部分（参考软件，Reference Software）：提供了用于演示功能和说明本标准其他部分功能的软件。

第 7 部分（一致性测试，Conformance Testing）：描述测试一个编码码流是否符合 MPEG—7 码流的方法。

2. MPEG—7 中的关键技术

为了对多媒体内容进行标准化的描述，MPEG—7 定义了一系列的标准化工具和方法。

- ① 描述符 D（Descriptors），对实体特征进行描述。
- ② 描述方案 DS（Description Schemes），规定了描述方案中各元素之间的结构和语义关系，这些元素既可以是描述符也可以是描述方案。
- ③ 描述定义语言 DDL（Description Definition Language），规定描述方案和描述符的语言，描述定义语言允许对现有的描述方案进行修改和扩展。

描述定义语言（DDL）、描述方案（DS）和描述符（D）之间的关系如图 5.22 所示。

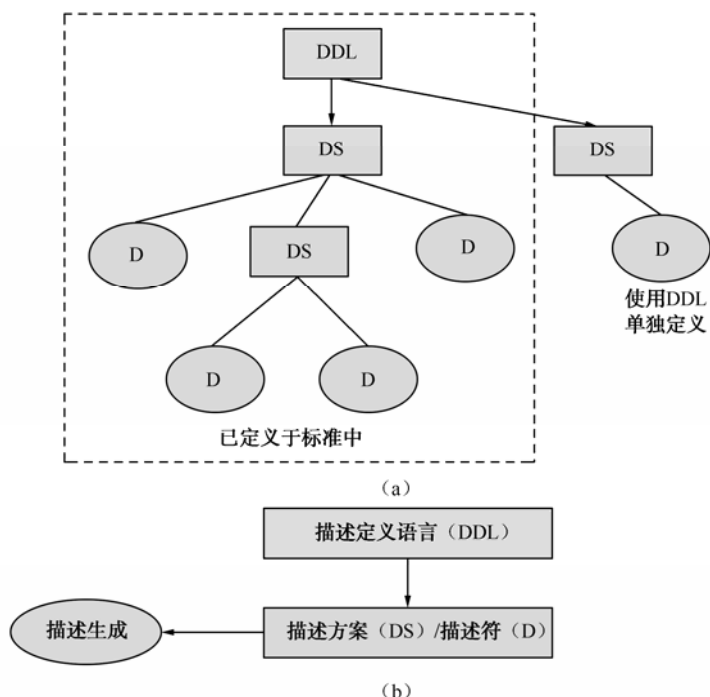


图 5.22 描述定义语言、描述方案和描述符之间的关系

从图 5.22 中可以看出描述定义语言 (DDL) 构成了整个 MPEG—7 内容的核心部分, 提供了建立描述方案和描述符的机制, 而描述方案和描述符则是描述生成的基础。MPEG—7 可以描述的多媒体对象包括静止图像、图形、三维模型、音频、语音、视频等, 以及如何组合这些资料的信息 (如 MPEG—4 中的场景描述信息)。

MPEG—7 可独立于其他 MPEG 标准使用, 但 MPEG—4 中所定义的音频、视频对象的描述适用于 MPEG—7, 这种描述是分类的基础。另外, 可以利用 MPEG—7 的描述来增强其他 MPEG 标准的功能。

MPEG—7 的应用领域包括: 数字图书馆 (Digital Library), 如图像目录、音乐词典等; 多媒体目录服务 (Multimedia Directory Services), 如黄页 (Yellow Pages); 广播媒体的选择, 如无线电频道、TV 频道等; 多媒体编辑, 如个人电子新闻服务、多媒体创作等。潜在的应用领域包括教育、娱乐、新闻、旅游、医疗、购物等。

5.3.5 MPEG—21

目前, 多媒体技术得到迅猛发展, 但多媒体信息交流仍有许多不便之处, 如不同网络之间的障碍、知识产权得不到有效保护等; 另外, 不同的多媒体信息、网络、设备、协议和标准、分布在不同的地点等都使用户不能以统一的方式进行多媒体信息交互。如何通过一个综合标准来对上述不便之处加以协调, 使多媒体业务畅通无阻呢? 为了解决这些问题, MPEG 从 2000 年 6 月开始着手制定 21 世纪多媒体应用的标准化技术: MPEG—21 “Multimedia Framework”。MPEG—21 是一个支持通过异构网络和设备, 使用户透明而广泛地使用多媒体资源的标准, 其目标是建立一个交互的多媒体框架。

1. MPEG—21 的组成与基本概念

MPEG—21 将形成一个开放的用于多媒体传送和消费的框架，是内容的创建者和消费者都关注的焦点。这一开放的框架既为内容的创建者和服务的提供者给予平等的机遇，又方便了内容的消费者以交互的方式存取大量不同的内容。MPEG—21 是由通过数字项而相互作用的用户组成的。MPEG—21 的行为主体是用户，规范的动作是“使用”，作用的客体是数字项，在 MPEG—21 作用的过程中形成了多媒体内容传送链和价值传送链。

以下是几个基本概念。

(1) 用户 (User)

MPEG—21 中的用户是一个广义的概念，可以是世界各地的个人、消费者、团体、组织、公司、政府及其他标准化组织和主体。从技术上讲，MPEG—21 认为内容的提供者和消费者同样都是 MPEG—21 的用户。用户可以以各种方式“使用”内容，包括对内容进行发布、递送和消费等。

(2) 使用 (Use)

MPEG—21 提供了一个多媒体框架供不同用户之间进行以数字信息为目标的相互作用，相互作用所涉及的各方均为用户。这里所指的相互作用即为对内容的“使用”，它包括内容创建、内容提供、内容存档、内容定级、内容增强和递送、内容聚集、内容传输、内容发表、内容零售、内容消费、内容提交、内容管制，以及以上各项交易的简化和以上各项交易的管制等。

(3) 数字项 (Digital Item)

MPEG—21 的数字项是一个结构化的数字对象，是按标准进行表达、标记并带有描述性的数据。数字项是数字资源及其相关内容（图像、Java Applet、数据文件、音频、视频等）的集合。它是 MPEG—21 框架中传送和交易的最基本单元，是用户之间发生关系的目的。

(4) 多媒体内容传送链

包含内容创建、生产、传送和消费等诸多环节。为了支持上述功能，应进行内容标志、内容描述、内容管理和内容保护；内容的传送和递送是在异构终端和网络上进行的，在此过程中，将发生事件并且产生事件报告。事件报告包括可靠的传送、个人数据的管理、用户隐私的优先考虑和对交易的管理。伴随着内容传送链的是抽象的价值传送链，表现在内容的传过程中其价值也可能发生变化或转移。

2. MPEG—21 中的关键技术

MPEG—21 包括多项技术，其中最主要的是以下几项。

- ① 数字项声明：是用于声明数字项的一种统一而灵活的抽象和互操作机制。
- ② 内容表示：解决如何表示媒体资源问题，使得内容可以被无“缝”地传送和消费。
- ③ 数字项的标志和描述：这是对不同自然属性、类型和粒度的数字项进行统一标记和描述的结构。
- ④ 内容的管理和使用：是一种接口和协议，用于保证通过内容分配和消费价值链来创建、制作、存储、传送和使用内容。
- ⑤ 知识产权管理和保护：用于确保内容在通过网络和设备时得到持久和可靠的保护。
- ⑥ 终端和网络：它们提供可交互和透明的通过网络和终端设备存取内容的能力。
- ⑦ 事件报告：是多媒体框架和用户之间的一种法则和接口，利用它可以使用户准确地了解在框架中发生的所有可报告事件的性能。

利用以上技术，用户的所有的工作可以概括为对内容的操作，如创作、提供、获取、评价、改进/传送、聚集、组织内容等。例如，在使用“音乐编辑”数字项时，不但可以编辑音乐，还可以穿插照片、视频、动画或图形、抒情诗、MIDI 文件等。又如你想制作自己所偏爱的乐曲 CD，首先可以通过网络收集乐曲。利用 MPEG—21 的数字项声明、标记和描述工具，用曲名、演奏者、歌词或旋律等不同方法来描述你所要查找的内容。在支持 MPEG—21 标记和描述的网络搜索引擎帮助下，可以收集到分布在世界各地异构网络上的数字化乐曲，再使用支持 MPEG—21 内容表现方法的媒体播放器来试听。当选定乐曲之后，便可以通过网络的电子支付手段购买乐曲的某种合适的使用权。在这一过程中传递的信用卡账号、密码等信息是保密而安全的。完成该过程之后，就可以将乐曲刻录到光盘上了。此外，还可以在网将乐曲所在的 URL（统一资源定位）路径，以及对它们的注解、点评等信息保存为一个工程文件，再从网上收集一些相关的图片甚至视频片段，加到这个 CD 工程文件中，然后将这个文件向其他人开放，就可以与其他人一起分享喜欢的音乐了。

3. 应用前景

当今世界正迈入数字化、网络化和全球一体化的信息时代，多媒体信息技术和互联网技术是促进社会全面实现信息化的关键技术。从 MPEG—1，MPEG—2，MPEG—4 到 MPEG—7，MPEG 标准正在向整个多媒体领域扩展，MPEG—21 就是一个有关多媒体框架及其综合应用的全新的多媒体框架标准。MPEG—21 远远超出了一个统一的活动图像压缩标准的范畴，必将对多媒体信息技术的广泛应用产生深远的影响。MPEG—21 多媒体框架标准将会为多媒体信息的用户提供综合统一的、高效集成的和透明交互的电子交易和使用环境，能够解决如何获取、如何传送各种不同类型的多媒体信息，以及如何进行管理、各种权利的保护、非授权存取和修改的保护等问题，为用户提供透明的和完全个性化的多媒体信息服务。MPEG—21 必将在多媒体信息服务和电子商务活动中发挥空前的重要作用。

5.4 H.26x

H.26x 是 ITU—T（国际电信联盟）及其前身 CCITT（国际电报电话咨询委员会）研究和制定的一系列视频编码的国际标准。其中，应用最为广泛的就是 H.261，H.262，H.263 和 H.264 这 4 个协议。虽然 H.26x 仅仅是视频编码的标准，而 MPEG—x 既包括视频编码标准，也包括音频编码标准和视音频同步标准。但是，二者之间却有着紧密的联系，因为从 MPEG—2 的制定开始，国际标准化组织（ISO）和国际电工委员会（IEC）就与 ITU—T 展开了合作，在一些 MPEG 标准中，H.26x 就是 MPEG—x 视频部分的重要组成。H.261 产生于 20 世纪 90 年代，可以说是视频编码的老前辈，如今已经逐渐退出历史舞台。H.262 是 MPEG—2 的视频部分，由于 MPEG—2 的应用十分广泛，因此 H.262 目前仍然是最重要的视频编码之一。H.263 是目前视频会议所采用的主流编码，在视频会议领域占有绝对的市场优势。H.264 是最近几年才刚刚出现的新的视频压缩标准，属于 MPEG—4 的第 10 部分。在相同的图像质量的情况下，H.264 有更高的压缩率，是一种很有市场潜力的视频压缩标准。

5.4.1 H.261

H.261 也称 $p \times 64$ ，可以说是视频编码的先驱者。H.261 标准设计的出发点是用于在 ISDN

(Integrated Services Digital Network, 综合业务数字网) 的通道上进行电视电话和电视会议。这实际上是属于在电话线路上传输非电话信号的“古老”问题。

考虑到 ISDN 的传输码率以 64kbps 为单位, 因此 H.261 以 $p \times 64\text{kbps}$ ($p=1 \sim 30$) 作为 H.261 的标准码率。它支持 QCIF (Quarter CIF, 1/4 通用中间格式) ($p=1, 2$)、通用中间格式 (Common Intermediate Format, CIF) ($p>2$) 两种图像输入格式。

H.261 标准主要解决以下问题。

① 编码算法问题: 确立了一种合理的、保证图像质量且为各国图像编码专家所公认的统一的算法。算法必须能够实时操作, 解码延时要短。

② 与 PCM 标准兼容的问题: 随着各种数字通信的发展, PCM 编码逐渐显得落后了。但在电话线路上传输非电话信号就必须考虑新的编码与 PCM 标准兼容的问题。H.261 解决这一问题的方法就是让编码器以 64~1920kbps 的工作速率去覆盖 N-ISDN 的通道。

③ 解决电视制式不同的问题: 为了使同一标准能用于 PAL (625) 和 NTSC (525) 两种电视制式系统, H.261 源编码采用了通用中间格式——CIF 格式 (参见 4.2.1 节)。所以对于 H.261 来说, 输入/输出必须先转换到 CIF 或 QCIF 格式后再进行源编码。

如图 5.23 所示是 H.261 编码/解码器的原理结构图。从图中可以看出, H.261 的编码过程分为 3 个部分, 信源编码器、图像复用编码器和发送编码器。信源编码器, 其主要任务是对图像信号进行压缩, 图像复用编码器的任务是将每帧图像数据分成图像 (Picture)、像块组 (Group Of Blocks, GOB)、宏块 (MacroBlock, MB) 和像块 (Block) 4 个层次的数据结构 (按照 CIF 格式, 每帧 CIF 图像包含 12 个 GOB, 每个 GOB 包含 33 个 MB, 每个 MB 包含 4 个亮度数据块和各 1 个 Cb, Cr 色度块, 每个 Block 包含 8 像素 \times 8 像素), 以便在各层次中插入必要的辅助数据。发送缓冲器的存储量是按照使用位率 $p \times 64\text{kbps}$, 加上固定富余量后确定的。发送编码器主要功能为插入正向纠错码, 以便接收终端的解码器能检测和纠正错误的码字。编码控制器, 除了控制量化步长外, 还控制编码模式, 即控制帧内编码或帧间编码, 这一操作是在信源编码器中进行的。

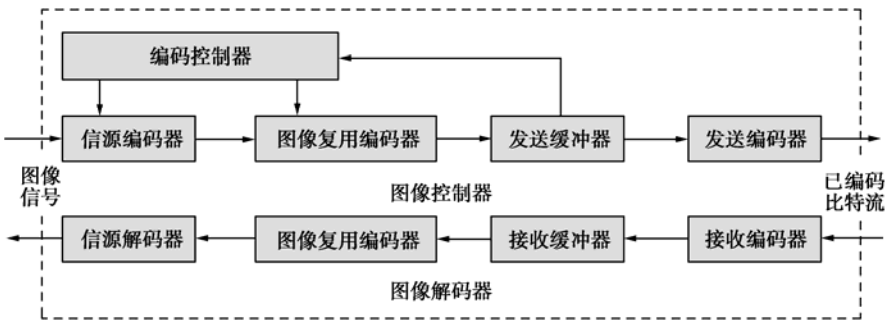


图 5.23 H.261 编码/解码器的原理结构图

从图像压缩的角度来看, H.261 主要采用了以下的压缩算法:

- ① 利用二维 DCT 减少图像的空间域的冗余度;
- ② 利用运动补偿预测减少图像的时间域的冗余度;
- ③ 利用视觉加权量化减少图像的“灰度域”的冗余度;
- ④ 利用熵编码来减少图像的“频率域”的冗余度。

5.4.2 H.262

ITU—T 于 1990 年成立了“ATM 视频编码专家组”，负责制定适用于 B-ISDN（Broadband Integrated Services Digital Network：宽带综合业务数字网）的信道 ATM 编码传输标准。该专家组于 1993 年 11 月与 ISO 的 MPEG 专家组联合提出了 H.262 建议草案，这一草案最终发展成为 H.262 标准，它也就是 MPEG—2 的视频部分。

H.262 标准是用于数字存储介质和数字视频通信中图像信息的编码表示和解码规定的。该标准向下兼容，能够在很宽的范围内对不同分辨率和不同输出比特的图像信号进行有效的压缩。

5.4.3 H.263

H.263 也是 ITU 制定的低比特率视频信号压缩标准，可以应用于 PSTN（Public Switching Telephone Network，公用电话交换网）和移动通信网。H.263 标准目前已经超过 H.261，在视频会议的编解码领域中占主导和支配地位（尽管 MPEG—4 的性能明显优于 H.263，但 MPEG—4 的应用还处在起步阶段）。在任何一种比特率的情况下，H.263 的性能都优于 H.261。而在相同的视觉效果下，H.263 又有比较高的压缩比。

H.263 标准制定了一种编码的表示方法，这种方法可以用来对低比特率视频设备构成的运动图像进行压缩。H.263 对于视频源编码法则的配置是基于 H.261 的建议，同时使用内插图片预测法，用临时的冗余和残余信号的变换编码来减少空间冗余。源编码器可以分为 5 种标准视频源格式：Sub-QCIF，QCIF，CIF，4CIF 和 16CIF（H.261 所支持的输入格式只有 QCIF 和 CIF）。源编码器也可以使用客户的其他视频格式。

H.263 编码/解码器的原理结构与 H.261 十分相似。H.263 的解码器具有运动补偿能力，而 H.263 的编码器可以选择是否加入运动补偿能力。H.263 标准是在 1996 年获得批准的，而在 1998 年和 2000 年，ITU—T 又对它进行了两次重要的修改和补充，于是产生了 H.263+ 和 H.263++。与原始的 H.263 相比，它们的功能更加强大。

总的来说，H.261，H.263 的压缩对象都是低比特率视频信号，而且 H.261~H.263 的压缩比还是不够高。这里所说的不够高是针对互联网传输而言的。对于在保持高清晰度的同时实现低比特率应用的需求，这 3 个标准都还无法满足要求。

5.4.4 H.264

MPEG—4（这里特指视频部分）和 H.263 这两种视频压缩标准大约都产生于 1995 年，而此后，动态图像专家组（MPEG）和视频编码专家组（VCEG）则仍然力图建立一个性能明显超过 MPEG—4 和 H.263 的新标准。所谓更高的性能包括更高的压缩比，保持高清晰的画质，以及低比特率视频流的应用。对于这样一个新的标准——高级视频编码（AVC）的研究可以一直追溯到 1995 年。

在 1995 年完成了 H.263 标准的最初版本后，ITU—T 的 VCEG 在两个领域开始了进一步的工作。

一项是短期的工作，目的是加强 H.263 的性能，这项工作的最终成果就是形成了 H.263+ 和 H.263++。

另一项是长期的工作，目的是开发一种低比特率视频应用的视频通信标准，这项工作导致了 H.26L 标准草案的产生。这是一个压缩比高于以前任何一种 ITU—T 标准的新标准。2001

年, ISO 的动态图像专家组认识到了 H.26L 标准的潜在价值, 因此与 VCEG 联合组成了联合视频小组 (JVT)。这个小组的主要工作就是把 H.26L 草案模型发展成为完全的国际标准。最终, 在 2002 年, 两个内容完全相同的标准产生了, 一个是 ISO MPEG—4 第 10 部分, 另一个是 ITU—T H.264 标准 (H.26L 在 ITU—T 的文件编号就是 H.264)。这一新标准的官方称谓是“高级视频编码 (AVC)”, 但 MPEG—4 Part 10 和 H.264 则更加广为人知。

由 ISO/IEC 和 ITU—T 制定的, 在 H.264 之前的视频标准并没有完全解决适应不同速率应用的问题和同时开会的需求。H.264 视频编码标准则在这些方面有了明显的进步, 它可以在几乎不降低画面质量的情况下, 将传输速率降低到 MPEG—2 (H.262) 的一半。制定 H.264 视频编码标准的目的在于改进编码效率和容错率, 以及能够更好地适应网络。

H.264 标准适用于多种应用软件和传输网络, 能够在多种实际应用中被使用, 可以达到不同的增强效率, 适应不同范围的比特率和清晰度。不过, 到目前为止 H.264 的商用还很少。

习 题 5

1. 填空题

- (1) 多媒体数据压缩方法根据不同的依据可产生不同的分类, 根据信息有无损失可分为: _____、_____。
- (2) 数据压缩技术的理论基础是_____, 解决的主要问题是_____、_____。
- (3) JPEG2000 采用以_____变换为主的多分辨率编码方式。
- (4) JPEG2000 支持_____编码, 能够处理图像中只有一小块区域对用户是有用的, 对这些区域采用低压缩比, 而其他区域采用高压缩比的情况。
- (5) 最近几年流行起来的所谓“MP3 音乐”就是一种采用_____层 3 编码的高质量数字音乐。
- (6) _____是 ISO 公布的“超低比特率活动图像和语音压缩标准”, 是针对低比特率下的视频、音频编码, 更加注重多媒体系统的交互性和灵活性, 引入了 AV 对象 (Audio/Visual Objects), 使得更多的交互操作成为可能。
- (7) _____是目前视频会议所采用的主流编码, 在视频会议领域占有绝对的市场优势。
- (8) H.26x 仅仅包括视频编码的标准, 而 MPEG—x 既包括视频编码标准, 也包括_____标准和_____标准。

2. 简答题

- (1) 为什么要进行数据压缩? 图像数据为什么可以压缩?
- (2) 压缩编码技术的发展共经历了几代? 简述每一代编码技术的主要特点。
- (3) 多媒体数据压缩性能的评价标准有哪些?
- (4) JPEG 压缩编码算法的主要步骤是什么? 哪些步骤对图像的质量是有损的? 哪些步骤对图像的质量是无损的?
- (5) JPEG2000 与 JPEG 相比有哪些优点?

- (6) 在 MPEG—1 编码码流中图像帧的传输顺序和显示顺序是一样的吗?
- (7) 什么是 MPEG 的运动补偿? 其目的是什么?
- (8) 什么是 MPEG 的宏块?
- (9) MPEG—1, MPEG—2, MPEG—4, MPEG—7, MPEG—21 的主要目标是什么?
- (10) H.261, H.262, H.263, H.264 的主要目标是什么?
- (11) H.26x 与 MPEG—x 有什么联系?

第6章 数字动画

动画技术的广泛应用，使我们的世界发生了翻天覆地的变化。在过去几十年里，计算机动画一直是人们研究的热点。随着计算机图形学和硬件技术的高速发展，人们已经可以用计算机生成高质量的图像，促使计算机动画技术飞速地发展起来。计算机动画的应用领域十分广泛，包括影视作品制作、科学研究、视觉模拟、电子游戏、工业设计、教学训练、军事仿真、过程控制、平面绘画、建筑设计等各个方面，使人们充分地体验到计算机动画高超技术的魅力。

6.1 动画基本概念

计算机动画是在传统动画的基础上，使用计算机图形图像技术而迅速发展起来的一门高新技术。动画使得多媒体信息更加生动，富于表现。从广义上看，数字图形图像的运动显示效果都可以称为动画。

6.1.1 动画的视觉原理

当我们观看电影、电视或动画片时，画面中的人物和场景是连续、流畅和自然的。但当仔细观看一段电影或动画胶片时，看到的画面却一点也不连续。只有以一定的速率把胶片投影到银幕上才能有运动的视觉效果，这种现象是由视觉残留造成的。动画和电影正是利用了人眼这一视觉残留特性。一秒钟呈现给眼睛图像的数量决定了景物的“闪烁率”。当眼睛能够测出每一图像帧时，便出现抖动（Flicker），这是因为帧与帧之间的时间间隔太长。实验证明，如果动画或电影的画面刷新率为 24fps 左右，则人眼看到的是连续的画面效果。但是，每秒 24 帧的刷新率仍会使人眼感到画面的闪烁，要消除闪烁感画面刷新率还要提高一倍。因此，每秒 24 帧的速率是电影放映的标准，它能最有效地使运动的画面连续流畅。但是，在电影的放映过程中有一个不透明的遮挡板每秒遮挡 24 次，因此电影画面的刷新率实际上是每秒 48 次。这样就能有效地消除闪烁，同时又节省了一倍的胶片。

在我国电视画面刷新率是 25fps。电视中没有遮挡板，为了降低闪烁现象，达到电影画面的同样效果，采用了交替传输的技术。在传输一帧画面时，首先是将画面的奇数行显示到屏幕上（组成一个奇数场），然后将画面的偶数行显示到屏幕上（组成一个偶数场），两场画面形成一帧。交错扫描相当于遮挡板的作用，从而减少了闪烁，提高了收视效果。在逐行扫描的计算机显示器中，画面刷新率还要高，一般在 75fps 以上。

6.1.2 计算机动画的概念

计算机动画技术是利用计算机产生运动图像的技术，具体说，就是利用计算机产生和处理帧画面上的图形和图像，交互地进行图形和图像的编辑、润色和声效结合，并将连续的画面实时演播的计算机技术。计算机动画的原理与传统动画基本相同，只是在传统动画的基础

上把计算机技术用于动画的处理和应用，并可以达到传统动画所达不到的效果。由于采用数字处理方式，动画的运动效果、画面色调、纹理、光影效果等可以不断改变，输出方式也多种多样。

早期，计算机曾用于控制动画机械设备，减轻了人工调整摄影机位置和校正动画摄制台的劳动强度。动画师给每个镜头输入一大堆指令，包括摄影机的高度、摄制台的 x , y 坐标值等，也包括渐显、渐隐、推拉、摇移等参数。计算机自动计算拍摄各帧的参数，并控制各个部件协调动作。计算机可以灵活、准确地控制镜头和控制台的运动，从而产生各种摄影的特效效果。

随着计算机技术（主要是计算机图形技术）的迅速发展，计算机在动画领域中的应用不断扩大，计算机动画的内涵也在不断扩大。计算机动画发展至今，主要分为两个阶段，即计算机辅助动画（Computer-Assisted Animation），也称为“二维动画”，以及计算机生成动画（Computer-Generated Animation），也称为“三维动画”。

目前计算机动画已经发展成一个多种学科和技术的综合领域，它以计算机图形学，特别是实体造型和真实感显示技术（消隐、光照模型、表面质感等）为基础，涉及图像处理技术、运动控制原理、视频技术、艺术，甚至于视觉心理学、生物学、机器人学、人工智能等领域，它以其自身的特点而逐渐成为一门独立的学科。

6.1.3 计算机动画的分类

计算机动画的分类方法有多种，按不同的方法有不同的分类。根据运动的控制方式可将计算机动画分为实时（Real-Time）动画和逐帧（Frame-By-Frame）动画两种。根据视觉空间的不同，计算机动画又可分为二维动画与三维动画。

1. 实时动画

实时动画也称为算法动画，它是采用各种算法来实现运动物体的运动控制的。在实时动画中，计算机一边计算一边显示就能产生动画效果，而在逐帧动画中，计算机生成每一帧画面需要一定的时间，为了达到动画效果必须待所有的画面生成后集中在一起播放才行。游戏软件以实时动画居多。电子游戏机的运动画面就是一种实时动画。在操作游戏时，人与机器之间的作用完全是实时、快速的。实时动画一般不必记录在磁盘上，观看时可以从显示器上直接实时显示出来。

在实时动画中，计算机对输入的数据进行快速处理，并在人们观察不到的时间内将结果显示出来。这种响应时间约在 $0.01 \sim 0.1s$ 之间，否则屏幕上的画面就有不连续的感觉。实时动画的响应时间与许多因素有关，例如，主机的工作速度，图形计算是使用硬件还是软件，所描述的景物是简单还是复杂，生成的图像是大还是小等。随着计算机技术的发展，不久的将来会出现实时的高分辨率、高质量的三维动画。

2. 逐帧动画

逐帧动画也称为帧动画或关键帧动画，也就是通过一帧一帧显示动画的图像序列来实现运动的效果，如图 6.1 所示。目前，三维动画一般采用逐帧生成画面的方法，有时生成一帧要几分钟乃至几个小时，生成的画面记录在磁盘上后再连续播放。制作者要有一定的制作经验，因为任何非实时的画面给人们的节奏感与实际播放时的感觉是迥然不同的。当因为错觉

而产生意料不到的视觉效果时，只有回过头来重新调整时间分配和节奏变化，才能获得满意的效果。

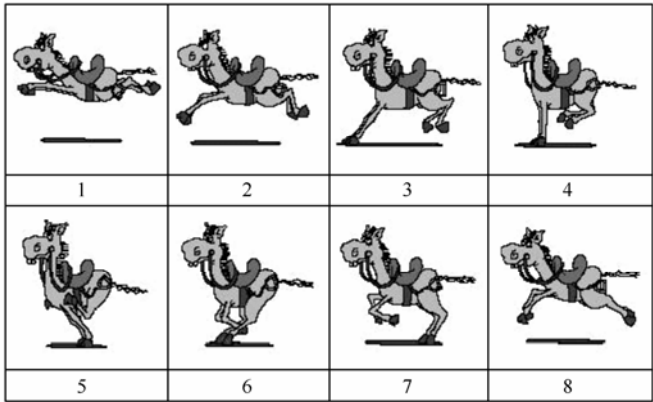


图 6.1 逐帧动画

关键帧技术是计算机动画中最基本并且运用最广泛的方法。关键帧技术来源于传统的动画制作。出现在动画片中的一段连续画面实际上是由一系列静止的画面来表现的，制作过程中并不需要逐帧绘制，只需从这些静止画面中选出少数几帧加以绘制。被选出的画面一般都出现在动作变化的转折点处，对这段连续动作起着关键的控制作用，因此称为关键帧（Key Frame）。在绘制出关键帧之后，再根据关键帧插入中间画面，就完成了动画制作。早期计算机动画模仿传统的动画生成方法，由计算机对关键帧进行插值，因此称为关键帧动画。

在二维、三维计算机动画中，中间帧的生成由计算机来完成，插值代替了设计中间帧的动画师。所有影响画面图像的参数都可成为关键帧的参数，如位置、旋转角、纹理的参数等。计算机插值有两种方式：线性插值和非线性插值，分别采用线性插值算法与非线性插值算法实现。关键帧技术通过对运动参数插值来实现对动画的运动控制，如物体的位置、方向、颜色等的变化，也可以对多个运动参数进行组合插值。我们在网上所见到的大多数 Flash 动画都是此类动画。

例如，在变形动画中（变形指景物的形体变化，它是使一幅图像在 1~2s 内逐步变化到另一幅完全不同图像的处理方法，如在电影《终结者 II》中机械杀手 T-1000 由液体变为金属人，由金属人变为影片中的其他角色（由 Alias 软件制作）），中间帧的生成是一种较复杂的图像处理，需要对各像素点的颜色、位置做变换。变形的起始图像和结束图像分别为两幅关键帧，从起始形状变化到结束形状的关键在于自动地生成中间形状，即自动生成中间帧。如图 6.2（a）所示，起始关键帧（一只鞋）和结束关键帧（一辆车）为预先处理好的相同幅面的图像。起始帧和结束帧上分别定义相同数量的若干组对应点 P_n 和 Q_n ，并使其按点顺序一一对应，这种对应关系确定了起始关键帧上一个点变化到结束关键帧的位置，如图 6.2（b）所示。通过变形算法生成由起始帧到结束帧的若干中间帧，连续播放即可构成由一只鞋变到一辆车的动画。

3. 二维动画

在传统卡通动画中，很多重复劳动可以借助计算机来完成。给出关键帧之间的插值规则，计算机就能进行中间帧的计算。实际上，在两个关键帧中隐含着很多信息，不具体提供这些

信息，计算机就很难进行计算，生成中间画面。

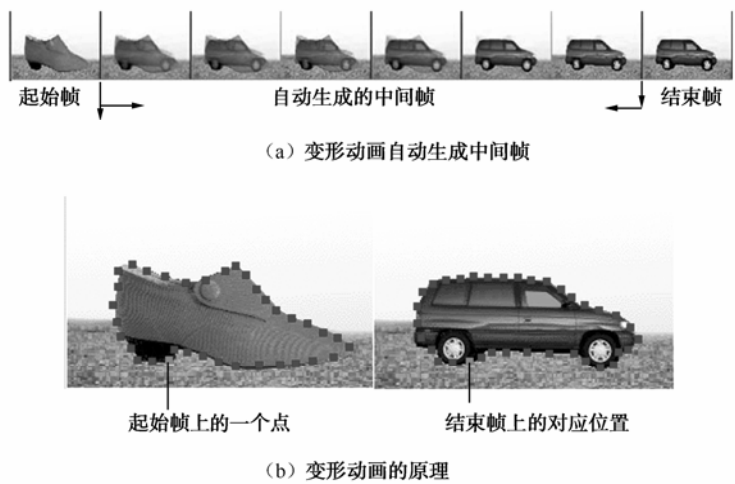


图 6.2 变形动画的中间帧和原理

一个比较简捷的办法是，将事先手工制作的动画逐帧输入计算机，由计算机帮助完成描线上色的工作，并且用计算机控制完成或记录工作。

二维动画是对手工传统动画的一个改进，它与手工动画相比有许多优越性。使用计算机进行描线上色工作非常方便，操作简单，颜色一致，绝对没有颜料的开裂、胶片闪光等问题。它的界线非常准确，不需晾干，不会串色，改色方便，制作工艺简单，不需通过胶片拍摄和冲印就能预演结果，发现问题可立即在计算机上修改。

在二维动画中，不仅具有模拟传统动画的制作功能，而且具有使用计算机所特有的功能，例如，计算机可以将图像复制、粘贴、翻转、放大、缩小、移位等。二维动画具有检查方便、保证质量、简化管理、提高制作效率、缩短制作周期等优点。

二维动画只能起辅助的作用，代替了手工动画中重复性强、劳动量大的那一部分工作，但代替不了人的创造性劳动，计算机不能根据剧本自动生成关键帧。目前，在二维动画中，关键帧必须由专业动画设计人员手工描绘，或者直接利用输入设备交互生成。

4. 三维动画

如果说二维动画对应于传统卡通动画的话，那么三维动画则对应于木偶动画。如同木偶动画中首先制作木偶、道具和景物一样，三维动画也需要首先建立角色、实物和景物的三维数据。接着，让这些角色和实物在三维空间里动起来：接近、远离、旋转、变形和变色等。再在计算机内部“加上”虚的摄影机，调好镜头，“打上”灯光，“贴上”材质，最后形成一系列栩栩如生的画面。三维动画之所以叫做计算机生成动画，是因为参加动画的对象不是简单由外部输入的，而是根据三维数据在计算机内部生成的，运动轨迹和动作的设计也是在三维空间中考虑的。

计算机动画真正具有生命力是由于三维动画的出现。我们所说的三维动画，不是简单地看起来具有立体感的动画，二维动画也可以看起来具有立体感；也不是说计算机制作的三维物体的画面就是真实物体的写照，事实上它的每一帧也是基于二维的像素点阵，只不过看起来像三维。例如，用三维处理来表现一个旋转的地球，首先要建立一个地球模型并把地图

贴满球面，然后使模型步进旋转，每次步进自动生成一帧动画画面，当然最后得到的动画仍然是二维的活动图像数据。如果在二维处理中，需要一帧帧地绘制球面变化画面，这样的处理难以自动进行。

二维与三维的主要区别在于图形中是否完全提供了深度信息。虽然二维中物体可以画得很像三维，但是当观察的视点改变时，画面必须由动画师另行画出。而三维动画由于物体有了三维深度的信息，仅描述一次就可以了，观察者可以从各个角度去观察，三维软件能自动计算光照与明暗程度。三维中也有二维功能，只是把二维看成三维的一个基础。

三维动画与二维相比，有一定的真实性，同时与真实物体相比又是虚拟的，这两者构成了三维动画的特性——虚拟现实。

6.1.4 计算机动画的制作

1. 二维动画的制作

在二维动画中，计算机的作用包括输入和编辑关键帧、计算和生成中间帧、定义和显示运动路径、交互式给画面上色、产生一些特技效果等。

二维计算机动画制作过程主要有以下几个步骤。

(1) 关键帧的产生

传统的动画创作，由美术师绘制关键的画面，再由美工使用关键画面描绘中间画面，最后逐一画面地拍照形成动画影片，其过程十分复杂。在计算机辅助动画设计中，关键帧及背景画面可以用摄像机、扫描仪、数字化仪实现数字化输入，也可以用相应软件直接绘制。动画软件都会提供各种工具，以方便绘图，还可以随时存储、检索、修改和删除任意画面。

(2) 中间帧的生成

利用计算机对关键帧进行插值计算，自动生成中间画面（如图 6.2（a）所示），这是计算机辅助动画的主要优点之一。这不仅大大提高了工作效率和质量，而且使画面精确、流畅。

(3) 分层制作合成

传统动画的一帧画面，是由多层透明胶片上的图画叠加合成的，这是保证质量、提高效率的一种方法，但制作中需要精确对位，而且受透光率的影响，透明胶片最多不超过 4 张。在动画软件中，也同样使用了分层的方法，但对位非常简单，层数从理论上说没有限制，对层的各种控制，像移动、旋转等，也非常容易。

(4) 着色

动画着色是非常重要的一个环节。动画软件一般都会提供许多绘画颜料效果，如画笔、调色板等。计算机动画辅助着色界线准确、修改方便，而且不因层数多少而影响颜色，速度快，且不需要为前后色彩的变化而头疼。

(5) 预演

在生成和制作特技效果之前，可以直接在计算机屏幕上演示一下草图或原画，检查动画过程中的动画和时限，以便及时发现问题并进行修改。

2. 三维动画的制作

在动画技术当中，最有魅力并应用最广的当然是三维动画。因为我们的世界是立体的，只有三维才能让我们感到更真实。二维动画可以看成三维动画的一个分支，它的制作难度及

对计算机性能的要求都远远低于三维动画。

计算机三维动画的制作过程主要有建模、编辑质材、贴图、灯光、动画编辑和渲染几个步骤。

（1）建模

建模就是利用三维软件创建物体和背景的三维模型，如人体模型（如图 6.3 所示）、飞机模型、建筑模型等。一般来说，先要绘出基本的几何形体，再将它们变成需要的形状，然后通过不同的方法将它们组合在一起，从而建立复杂的形体。

（2）编辑材质

就是按要求对模型的光滑度、反光度、透明度进行编辑。如玻璃的光滑和透明、木料的低反光度和不透明等都是在这一步实现的。如果经过这一步后直接渲染，我们就可以得到一些漂亮的单色物体，如玻璃器皿和金属物体。

（3）贴图

我们现实生活中的物体并不都是单色的物体，除了独特质感，现实物体的表面都有丰富的纹理和图像效果，如木纹、花纹等，这就是要给对象贴图的原因。在三维动画中要做得逼真，也要将这些元素做出来。如果直接在三维的模型上做出这种效果是难以实现的。所以一般都是将一幅或几幅平面的图像像贴纸一样贴到模型上，这就是贴图，如图 6.4 所示。

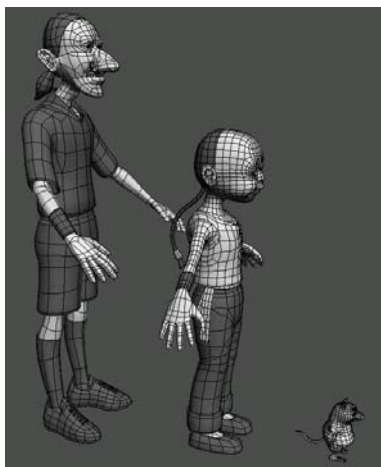


图 6.3 三维动画的人体模型



图 6.4 三维动画的贴图

（4）灯光

三维环境中最重要的是照明，灯光照明是最常用的方式。在场景中不同位置放上几个光源，从不同的角度用灯光照射物体，可烘托出不同的光照效果。灯光有主光和辅光之分，主光的任务是表现场景中的某些物体的照明效果，一般需给物体投影，辅光主要是辅助主光在场景中进行照明。

（5）动画编辑

做出来的模型是静态的物体，要使其运动起来就要经过动画编辑。动画编辑就是使各种造型运动起来，由于计算机有非常强的运算能力，制作人员所要做的就是定义关键帧，中间帧交给计算机去完成，这就使人们可做出与现实世界非常一致的动画。

（6）渲染

渲染一般指的就是把我们所做的模型、设置的灯光材质等各种对象综合到一起，制作成一个具有真实效果的图像文件。3D 建模和动画往往仅占全部动画制作过程中的一部分，大部分时间都花费在繁重的渲染工作中。渲染工作对处理器的处理性能有极强的依赖性。因此，为了获得更高的渲染性能，用户必须尽可能地使用更高性能和更多数量的处理器。

6.1.5 计算机动画的应用

计算机动画的应用十分广泛，可用于影视领域中的电影特技、动画片制作、片头制作、基于虚拟角色的电影制作等；还有电视广告制作、教育领域中的辅助教学、教育软件等；科技领域中的科学计算可视化、复杂系统工程中的动态模拟；视觉模拟领域中的作战模拟、军事训练和驾驶员训练模拟；此外还有娱乐业中的各种大型游戏软件，尤其是与虚拟现实技术相结合，将会创建各种幻想游乐园。今天，计算机动画已渗透到社会的许多方面，下面将介绍计算机动画在几个主要方面的应用情况。

1. 在电影工业中的应用

计算机动画使用最多的可能要数电影工业了。早在 20 世纪 60 年代，两位来自美国 Bell 实验室的科学家 Messrs.Zajac 和 Knowtion 就开始了这方面的尝试，后来由于计算机图形学方面的进步和一系列图形输出设备的推出，在电影界开始用计算机代替手工制作动画。据资料介绍，近年来所推出的影视作品中的动画和许多特技镜头，大都是计算机的杰作。看过《侏罗纪公园》这部电影的读者一定会对影片中那些栩栩如生的庞然大物——恐龙记忆犹新，如图 6.5 所示。它能和演员同处一个画面，并能将汽车掀翻。这个影片中的所有的动画镜头全是计算机制作的，其效果达到了以假乱真的程度。另外，星球大战也是一部许多人熟悉的科幻影片，在影片中出现的 X 机翼的战斗机，看上去和真实的模型没有任何区别。



图 6.5 《侏罗纪公园》中的恐龙

利用计算机动画制作电影的好处在于能让计算机控制物体的运动，无须重构每一步。这样便提高了真实感，并且降低了制作成本。然而用计算机制作动画也需较长的时间，动画的质量越高所需的时间越长，因为其中将涉及许多复杂的数学计算。这些数学公式能被用于处理景物和产生带有特殊效果的真实感的图像。时至今日，计算机图形学和计算机图形硬件的发展已取得很大的突破，一些厂家已相继推出了面向动画制作和图像处理的图形工作站。制作动画对大多数人来说已不再是一件难事。然而，好的动画设计毕竟还需要艺术天赋，尤其是用于影视艺术的动画。而对于一般的动画制作，今天的软件已能使大多数初学计算机的人

方便地制作动画，其过程基本上是自动的。

2. 在教育中的应用

计算机动画在教育领域中的应用有着光辉灿烂的未来。随着个人计算机的不断普及，将会有越来越多的课程利用计算机辅助教学，而在计算机辅助教学中，动画则是一种人们喜闻乐见的信息表示形式。例如，利用动画可以教幼儿识数，辨别上、下、左、右；利用动画可以演示一个物理定律，说明一个化学反应过程。目前，我国已有为数众多的计算机辅助教学软件用于幼儿园、小学、中学、大学乃至职业培训，如图 6.6 所示。

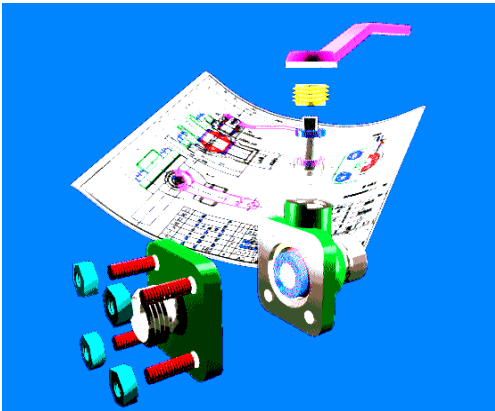


图 6.6 计算机辅助教学软件

在这些软件中，出现了大量的计算机动画，学习者可以自己操纵计算机，计算机按照人们输入的信息显示各种信息和动画的运动过程，这会极大地提高学习者的兴趣，巩固所学的知识。例如，有些化学实验的化学反应，需要一定的时间（有的长达几天），并且若操作不当还会发生爆炸、燃烧等危及人身安全的情况，同时，化学实验还需耗费大量的实验材料。而利用计算机动画模拟的化学实验，学生只需在计算机上选择所要做的实验，以及进行该实验的材料、步骤，计算机便会用动画动态地模拟实验的每一步过程，给出反馈信息和学生学习情况，使实验者从计算机屏幕上能够一目了然地获得实验数据。在教育各个层次与此相似的例子比比皆是，数不胜数。

3. 在科学研究中的应用

动画在科学研究中被大量用来模拟和仿真某些自然现象、物体的内部构造及其运动规律。在空间探测领域中，计算机动画被用来模拟飞行器或行星的运行轨道或太空中的某些自然现象。凡看过卫星发射电视转播的人应该都还记得，在卫星发射中心的控制室的大屏幕上能动态地画出卫星的运行轨道及所处的位置，使控制中心的工作人员一目了然，这便是计算机动画所起的作用。当卫星发射后，各种测量仪器将测量的卫星飞行数据源源不断地送往控制中心的计算机中，计算机再根据这些数据，准确、及时地在屏幕上画出卫星的飞行情况。

早在 1986 年 1 月，由美国国家航空航天局发射的先驱者和旅行者空间探测器的探测情况，就被喷气推进实验室的科学家们根据所接收的观察数据和太空的自然运动法则来动态地显示在了计算机屏幕上。美国国家航空航天局的科学家们能够直观地了解到太空行星特定轨道和太空中观察的景色，就好像科学家们自己乘坐探测器观察的那样。这个软件还允许选择

观察的视角，将观察点放在探测器的后面，这样可以看到探测器也可以看到行星。如图 6.7 所示为美国旅行者号火星车着陆火星表面行走的动态模拟图。



图 6.7 美国旅行者号火星车着陆火星表面行走的动态模拟图

在医学研究中，计算机动画能够帮助医生和研究者可视化地构造特定的器官和骨骼结构，分析病人的病症，好对症下药。如今像这些带有计算机动画功能的医疗设备在一些大的医院和医学研究机构已随处可见。

4. 在训练模拟中的应用

计算机动画也可用于训练模拟。例如，在运动员训练中，可以利用计算机帮助运动员改进他们的动作。如在一个运动员跑步时，计算机能根据捕获的图像数据，分析运动员训练时存在的问题，给出相应的训练建议和动作要求，其中，动作的要求也由计算机用动画产生，运动员可根据计算机的动画演示来进行动作训练，同样的思想可用于游泳、网球等。据资料介绍，采用这种辅助训练系统，对改正运动员不规范的动作，提高运动成绩有很大的帮助。

计算机动画技术在飞行模拟器的设计中起着非常重要的作用。该技术主要用来实时生成具有真实感的周围环境图像，如机场、山脉和云彩等。此时，飞行员驾驶舱的舷舱成为计算机屏幕，飞行员的飞行控制信息转化为数字信号直接输出到计算机程序，进而模拟飞机的各种飞行特征。飞行员可以模拟驾驶飞机进行起飞、着陆、转身等操作，如图 6.8 所示。



图 6.8 飞行模拟器

5. 在工程设计中的应用

计算机辅助设计（CAD）在如今的工程界已不再是一个新的名词了，在世界许多国家有大量的计算机用于工程设计，如今的 CAD 软件已做到设计完成后能动态地将设计结果用三维图形显示出来，如图 6.9 所示。例如，一个机械设计师，当为某一机器设计了一个部件后，计算机便可模拟这个部件的真实情况，能以不同的光洁度和不同的视角显示设计结果。如果是一组配套部件，还能够显示装配过程。



图 6.9 计算机辅助设计应用

6. 计算机动画在艺术和广告中的应用

计算机和艺术家相结合无疑会给艺术家的艺术创作提供极大的便利和许多艺术灵感。计算机的绘画软件能提供更多的色彩，提供使物体更具真实感的各种光照模型，并且用计算机画画、修改也极为方便。

在广告领域中，计算机动画是大有用武之地的，如今各类电视广告在各种节目中出现，而在这些广告中，有相当一部分是利用计算机动画来制作产生的。如今某些专用动画软件的功能是许多艺术家所望尘莫及的，而对使用者的要求却很低，只要略懂一点计算机就行了。计算机动画除了影视广告中的应用之外，在各类信息板、广告牌中也大量使用。如今当我们穿梭在繁华闹市或暂留在车站码头时，到处可见五颜六色的各类大型电子广告牌，而这些广告牌中显示的各种文字、图案、动画均是计算机的杰作。如图 6.10 所示为计算机制作的汽车广告。

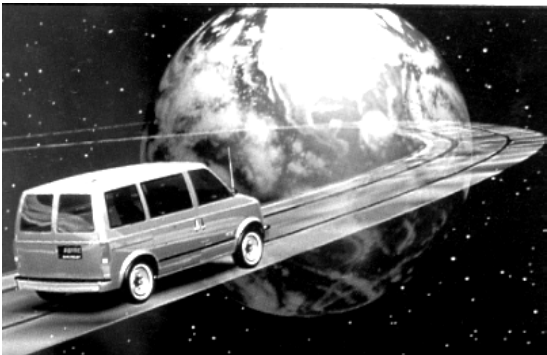


图 6.10 计算机制作的汽车广告

6.1.6 计算机动画的发展趋势

自从 Bell 实验室于 1963 年制作了第一部计算机动画片以来，计算机动画技术已有了很快的发展。计算机动画的研究涉及具有人的意识的虚拟角色的集成动画系统。研究内容不仅包含动画描述模型、动画语言、运动控制方法，还包括关键帧的生成技术、三维动画中的物体造型技术、动画的相关技术、动画的生成与绘制技术。总之研究内容涉及多种学科的知识、技术和方法，如动画、力学、机器人学、生物学、心理学和人工智能等。研究计算机动画所要实现的目标：能自动产生计算机生成的虚拟角色——人的自然行为；提高计算机动画运动的复杂性和真实性；应减少运动描述的复杂性，特别是可在任务级上进行运动的描述，从而解决制作复杂动画的很多难题。在不久的将来导演可在视频屏幕前，使用不同的命令来导演虚拟角色、灯光、舞台布置和摄像机。如果这一切能实时完成（前提为计算机硬件性能大大提高了，数据存储设备能提供足够的速度和容量，软件所能实现的效果不受计算机时间和空间的影响，且人们对人脑和人类行为的认识基本达到了一定的程度），那将像是在虚拟世界里导演一部真实的影片。那时把计算机动画、多媒体、人工智能等多种技术相结合将会完善虚拟环境技术，即动画技术的发展，会促进虚拟现实的进一步完善，实现智能的人-机交互界面，从而真正地实现虚拟环境系统。可以预见，不久的将来，虚拟现实能对人类的生活产生重大的、前所未有的影响。

6.2 GIF动画

GIF 是图形交换格式（Graphics Interchange Format）的英文缩写，是由 CompuServe 公司于 20 世纪 80 年代推出的一种高压缩比的彩色图像文件格式。GIF 文件可以制作动画，这是它突出的一个特点，该动画格式一直是 Internet 上应用最广泛的动画格式之一。GIF 采用了无损压缩方式，在不影响图像质量的情况下可生成很小的文件。该文件支持透明色，可使图像浮现在背景之上，缺点是只支持 256 色。

GIF 文件的制作与其他文件不太相同。首先，要在图像处理软件中做好 GIF 动画中的每一幅单帧画面，然后再用专门的制作 GIF 文件的软件把这些静止的画面连在一起，再定好帧与帧之间的时间间隔，最后再保存成 GIF 格式就可以了，如图 6.11 所示。制作 GIF 文件的软件也很多，比较常见的有 Animagic GIF、GIF Construction Set（简称 GIFCON）、Ulead GIF Animator（详见实验 T4.1）等。

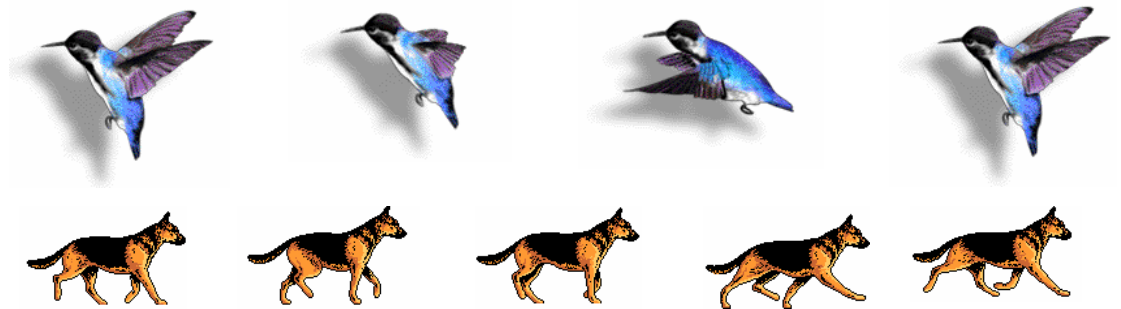


图 6.11 GIF 动画示例

6.3 Flash动画

现在，我们在浏览网页时，经常会看到引人入胜的动画效果，制作这种效果的工具就是Flash。在二维动画的软件中，Flash可以说是后起之秀，它已无可争议地成为最优秀的交互动画的制作工具，并迅速流行起来。Flash的动画效果不再是单纯的反复运动，而是可以在画面里进行菜单选择和操作，以及播放声音文件。Flash的出现，立即受到用户的欢迎，并迅速被应用到网络与多媒体课件设计中。

6.3.1 Flash动画的制作方式

Flash动画只包含两种基本的动画制作方式，即补间动画和逐帧动画。Flash生成的动画文件的扩展名默认为.flg和.swf，前者在Flash环境中运行，后者可以脱离Flash环境运行。

1. 补间动画

补间动画可用于创建随时间移动或更改的动画，例如，对象大小、形状、颜色、位置的变化等。在补间动画中，用户只需创建起始和结束两个关键帧，而中间的帧则由Flash通过计算自动生成。由于补间动画只保存帧之间更改的值，因此可以有效减小生成文件的大小。

补间动画分为补间动作动画和补间形状动画两种，其区别如下：

① 补间动作动画。在改变一个实例、组或文本块的位置、大小和旋转等属性时，可使用补间动作动画。使用补间动作动画还可以创建沿路径运动的动画。

② 补间形状动画。在改变一个矢量图形的形状、颜色、位置，或使一个矢量图形变为另一个矢量图形时，可使用补间形状动画。

2. 逐帧动画

逐帧动画是一种传统的动画形式，在逐帧动画中用户需要设置舞台中每一帧的内容。由于逐帧动画中Flash要保存每个帧上的内容，因此采用逐帧动画方式的文件容量通常要比采用补间动画的文件容量大。

逐帧动画模拟传统卡通片的逐帧绘制方法，不仅耗时，而且要求用户具有较高的绘图能力。补间动画则不然，由于所有中间帧均由工具自动完成，使不会绘画的用户也可轻松地作出形状和色彩逐渐变化、移动速度快慢随意的动画，动画文件的容量也较逐帧动画小得多，因而更适合于绘画水平不高的初学者使用。

Flash动画与其他动画的一个基本区别就是其具有交互性。所谓交互就是通过使用键盘、鼠标等工具，可以在作品各个部分跳转，使用户参与其中。

Flash是基于矢量的图形系统，各元素都是矢量，只要用少量的向量数据就可以描述一个复杂的对象，占用的存储空间只是位图的几千分之一，非常适合在网络上使用。同时，矢量图像可以做到真正的无级放大，这样，无论用户的浏览器使用多大的窗口，图像始终可以完全显示，并且不会降低画面质量。

Flash使用插件方式工作。用户只要安装一次插件，以后就可以快速启动并观看动画。由于Flash生成的动画一般都很小，所以调用的时候速度很快。

Flash通过使用矢量图形和流式播放技术克服了目前网络传输速度慢的缺点。基于矢量图

形的 Flash 动画尺寸可以随意调整缩放，并且文件很小，非常适合在网络上使用。流式技术允许用户在动画文件全部下载完之前播放已下载的部分。

6.3.2 Flash动画的制作软件

Flash 是美国 Macromedia 公司出品的矢量图形编辑和动画创作的软件，它与该公司的 Dreamweaver（网页设计）和 Fireworks（图像处理）组成了网页制作的“三剑客”。Flash MX（详见实验 T4.2）是 Flash 比较流行的版本，简单易学，用户可以很轻松地掌握，并制作出效果非凡的 Flash 动画。目前的最新版本是 Flash CS4，功能非常强大，不仅支持骨骼动画，还支持 3 维等等，其界面如图 6.12 所示。



图 6.12 Flash CS4 界面

6.4 三维动画制作

6.4.1 三维动画的原理

由于三维动画的表现形式更加直观，早期人们为了创作三维动画，不得不用木头、泥土或纸张等建立各种各样的三维模型，再设法使其运动。然而，在现实世界中建立一个三维模型需具有一定的专业技能，并且建立模型的过程是一件令人乏味的事，一个模型一旦建立，若要修改必须投入大量的劳动。人类为了方便地交流信息，更多的是将这些三维物体在一个平面上（如纸上）表示。如今即使一个最复杂的三维结构也能以二维形式表示出来，并且这种表示方式被大量用于工程设计和影视动画。

随着计算机技术的进步和计算机图形学的发展，特别是微型计算机的迅速普及，已有越来越多的人感受到用计算机制作三维模型和动画的优越性。首先获利的是工程设计和影视制作。如今设计工程师们能够利用计算机辅助设计（CAD）系统方便地建立设计模型，让计算机自动画出该模型的各种图纸，并能获得用其他物理模型都无法获得的视觉效果。例如，建筑设计师们能够在计算机上产生他们设计的建筑模型，他们能够“进入”计算机产生的房子里面，从居住者所希望的视角来观察。他们也能快速、容易地修改一个计算机产生的模型，并能为模型选择建筑材料。计算机三维动画也给影视业制作注入了新的活力。使用计算机人们能够较容易地创作各种动画角色和特技效果，采用现有的视频技术能使计算机动画产生的角色和许多著名影星同场演出，目前已出现了计算机“演员”，例如图 6.13 所示的《谁陷害了兔子罗杰》中的动画角色。

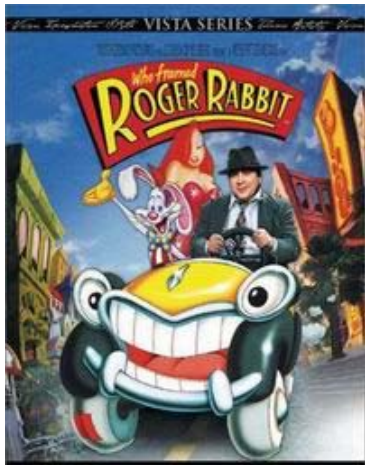


图 6.13 《谁陷害了兔子罗杰》中的动画角色

正如我们所知道的那样，动画是一种基于时间和空间的媒体，正由于这些，计算机三维动画能让我们自主地控制自己的信息空间。利用动画能够在几秒钟内有效地显示一个长的时间过程（比如土壤的分化），相反，为了便于理解一些瞬间即逝的事件发生过程，利用动画可减慢其发生过程，让它在 5~10s 内发生。

早期在计算机上建立三维动画是靠用某一种计算机语言编写程序来实现的，这需要有较高的计算机、数学和艺术素养。在计算机技术迅速发展的今天，对一般用户而言，没有必要从基础做起，因为如今在各类计算机上已有足够多的三维动画软件或工具供选择。用这些软件建立三维动画一般来说有 6 个基本的步骤，这对大多数软件包而言是共同的，无须考虑正在使用的计算机平台。这 6 个基本步骤是建模、编辑材质、贴图、灯光、动画编辑和渲染。

6.4.2 三维动画的制作软件

1. 3D Studio MAX

3D Studio MAX 是由美国 Autodesk 公司推出的基于个人计算机的三维动画设计和制作软件。3D Studio MAX 为专业的三维电影电视设计，同时兼顾交互游戏的设计及其他方面的应用。对于工程师来说，3DS MAX 在静态渲染、动态漫游、产品仿真及实现虚拟现实的过程中起着越来越大的作用。由于 3D Studio MAX 功能强大，并较好地适应了国内 PC 用户

众多的特点，因此被广泛运用于三维动画设计、影视广告设计、室内外装饰设计等领域。其主界面如图 6.14 所示。

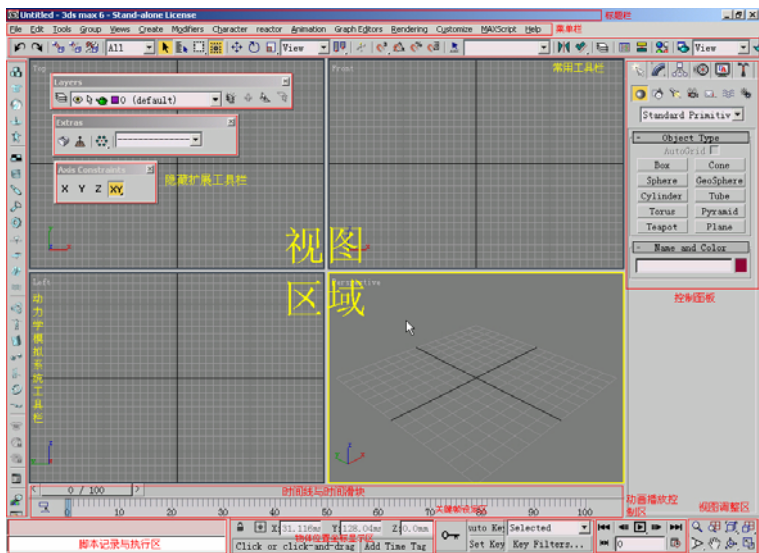


图 6.14 3DS MAX 6.0 主界面

2. Maya

Maya 是 Alias/Wavefront 公司在 1998 年推出的三维动画制作软件，虽然相对于其他（如 3DS 等）经典的三维制作软件来说，Maya 的历史并不长，但 Maya 凭借其强大的功能，友好的用户界面和丰富的视觉效果，一经推出就引起了动画和影视界的广泛关注。目前 Maya 已成为世界上最为优秀的三维动画的制作软件之一。

Maya 提供了适用于 Windows NT, Mac, Linux 等不同平台的版本，还可以在 SGI IRIX 操作系统上运行，广泛用于专业的影视广告、角色动画、电影特效特技等领域。Maya 具有功能完善、操作灵活、易学易用、制作效率极高、渲染真实感极强等特点。Maya 能极大地提高制作效率和品质，调节出仿真的角色动画，渲染出电影般的真实效果。Maya 集成了最先进的动画及数字效果技术，它不仅包括一般三维和视觉效果制作功能，而且还能与最先进的建模、数字化布料模拟、毛发渲染、运动匹配技术相结合来使用。

Maya 给三维设计者提供了优秀的制作工具来表达无比的创意，其主要特性如下。

- ① 采用节点框架，可即时修改和描述动画并能记忆制作过程。提供关键帧和程序动画的制作工具，可以迅速而容易地设定驱动键，提高工作效率。
- ② 用新的技术代替传统的关键帧创造复杂动画，可将多层动画路径复合成单一路径，并以简单曲线操控由运动捕捉器（Motion Capture）产生的密集曲线。
- ③ 提供理想的肌肉、皮肤和服饰动画制作模型工具，同时可生成自然景观。
- ④ 支持复杂的动态交互功能。
- ⑤ 在建造模型方面，提供完整的制作模型工具、变形工具箱、编织曲面。
- ⑥ 在上色方面，提供了选择式光学追踪法，模拟各种透镜及灯光效果，如闪光、云雾、逆光、眩光等。

Maya 是为影视创作应用而开发的，从《星球大战前传》、《透明人》、《黑客帝国》、《角

斗士》、《完美风暴》、《恐龙》(如图 6.15 所示)等到很多大片中的电脑特技镜头都是应用 Maya 完成的。除了影视方面的应用外, Maya 在三维动画制作、影视广告设计、多媒体制作及游戏制作领域都有很出色的表现。



图 6.15 利用 Maya 制作的三维动画片《恐龙》

逼真的角色动画、丰富的画笔,接近完美的毛发、衣服效果,不仅影视广告公司对 Maya 情有独钟,许多喜爱三维动画制作,并有志向影视电脑特技方向发展的朋友也为 Maya 的强大功能所吸引。

习 题 6

1. 填空题

- (1) 我们所看到的动画,实际上是由若干幅静止图片所组成的。之所以能有动的感觉,主要是由人类的_____生理现象所致。
- (2) 计算机动画的分类方法有多种,根据运动的控制方式可将计算机动画分为_____和_____两种。
- (3) _____技术是计算机动画中最基本并且运用最广泛的方法。
- (4) GIF 是_____的英文缩写,该动画格式一直是 Internet 上应用最广泛的动画格式之一。
- (5) Flash 生成的动画文件的扩展名默认为_____和_____。前者在 Flash 环境中运行,后者可以脱离 Flash 环境运行。
- (6) _____一般不必记录在磁盘上,观看时可以从显示器上直接实时显示出来。

2. 简答题

- (1) 什么是动画?什么是计算机动画?
- (2) 用计算机实现的动画最常见的可分为哪几类?
- (3) 简述二维动画的制作过程。
- (4) 简述三维动画的制作过程。
- (5) 列举常用的二维动画制作软件,并说明其各自的特点。
- (6) 列举常用的三维动画制作软件,并说明其各自的特点。

第7章 多媒体电子出版物与创作工具

电子出版物是计算机技术、多媒体技术、大容量光盘存储技术，以及网络技术等诸多技术领域综合发展的产物。电子出版物具有传统媒体无法替代的优点，它具有交互性强、存储信息量大、图文声并茂、易于交流与携带、利于共享等特点，因而得到了迅速的普及。目前，电子出版物在形式、种类、内容、规模和技术各方面都发展迅速，为信息的处理与传播提供了最有力的支持。

本章将介绍制作电子出版物所涉及的技术与环境，以及多媒体创作工具。

7.1 多媒体电子出版物

7.1.1 多媒体电子出版物的概念与特点

对于电子出版物，国家新闻出版总署有明确的定义：“电子出版物，系指以数字代码方式将图、文、声、像等信息存储在磁、光、电介质上，通过计算机或类似设备阅读使用，并可复制发行的大众传播媒体”。这个定义明确了电子出版物的几个重要特点：信息多媒体化，存储形式和载体多媒体化、数字化，阅读使用计算机化。

电子出版物在载体形式上主要可分为光盘出版物和网络出版物，光盘出版物现在已得到普遍发展，是电子出版物的主要形式。另外，电子出版物可用超文本格式制作成“无纸无盘”的网络联机版，通过互联网在线发布。

从内容和应用领域上分，电子出版物大体可分为电子图书、辞书手册、文档资料、报纸刊物、教育培训、娱乐游戏、宣传手册、信息咨询、电子简报等一些类型。事实上许多作品是多种类型的混合，难以明确划分其类型。

与传统媒介相比，电子出版物具有以下几个特点。

(1) 丰富的多媒体信息表现

电子出版物可以表现传统出版物中所无法表现的动态多媒体信息，如音频、视频、声音、动画等。

(2) 容量大、体积小

一张 CD-ROM 光盘可存放二十多卷的百科全书，或 1~2 年每天数十个版面的报纸，体积小，收藏管理方便。

(3) 交互能力与检索查询

借助于超文本技术和计算机的交互处理能力，可以对信息进行有效的组织，因而能方便快速地检索、查询所需的内容。还可以将光盘中的内容复制或打印。

(4) 制作高效、出版迅速

电子出版物的整个制作过程都需借助于计算机完成，处理方便、速度快、效率高，手工加工环节少，因而出版速度快，周期短。

（5）成本低、节省能源

以单位信息量计算，电子出版物比纸质出版物的加工成本少得多，耗用资源少，且发行速度快。

7.1.2 多媒体电子出版物的开发与制作

制作多媒体电子出版物，分两个阶段。首先是素材的获取与加工，其次是多媒体信息的编辑和创作。

基于此，多媒体电子出版物制作系统分为两大部分。一是素材获取与加工系统，包括文字录入系统、音频采集系统、图像扫描系统、图形绘制系统、动画创作系统和视频捕获系统等；二是多媒体编辑和创作系统，将零散、非连贯的各种素材信息整合到一起，进行多媒体编辑、创作，使其彼此之间按照有机的方式交互联系，具备良好的可读性。

多媒体电子出版物的开发与制作流程，一般分为 10 个阶段：选题策划，脚本创意，系统设计，素材积累，图文处理，动画制作，音、视频信号处理，多媒体编辑创作，样盘测试及批量制作发行。

7.2 多媒体创作工具

多媒体创作工具是指能够集成处理和统一管理文本、图形、图像、视频、动画、声音等多种媒体信息，使之能够根据用户的需要生成多媒体应用程序的编辑工具。多媒体创作工具又称为多媒体开发平台或多媒体创作工具。

近几年来，各类多媒体创作工具应运而生。多媒体创作系统向管理人员、教育工作者及专业设计者提供了一种途径，即无须成为程序设计专家就可以将不同的媒体融合在一起，成为制作电子出版物的大师，这个途径就是学习、掌握多媒体创作工具。它无须掌握深奥的计算机编程技术，面向广大非计算机专业人员，易于学习和使用，可大大缩短开发周期，这种开发方法已被大多数人所接受。

7.2.1 多媒体创作工具的特点

多媒体创作工具产生的初衷是为不懂编程的应用人员制作多媒体应用软件提供一种便利的工具。由于多媒体创作工具这种用户的“非专业性”，它必须具有概念清楚、界面简洁、操作简单、易用易学等特点。

目前，较为流行的多媒体创作工具普遍都具有以下这些特点。

1. 集成性的开发环境

多媒体创作工具可以随时利用其他软件编辑、制作的多媒体数据素材，并可对这些素材进行编辑处理，使之按照一定的要求，交互性地呈现。有些多媒体创作工具带有强大的文本、图形创作工具，可以编辑制作出多媒体应用软件需要的素材。多媒体应用软件编辑完成以后，可以利用多媒体创作工具生成脱离环境独立运行的可执行文件。

2. 交互性的、面向对象的操作环境

多媒体创作工具为创作人员提供了交互性的、面向对象的操作环境。对出现在多媒体应

用软件中的任意多媒体信息，创作人员可以根据需要，对它们进行编辑处理，如移动、延时、声音起止时间的设定、图形呈现与消隐特技等，这些控制方式都是全方位的。

3. 操作简便

多媒体创作工具的各项功能，无须通过编程实现。多媒体应用程序的创作人员只需根据需要，操作特定的功能键、鼠标等，即可实现对多媒体信息的编辑处理。因此，通过一定阶段的学习，一般人员都可以利用多媒体创作工具创作多媒体应用程序。这也为多媒体创作工具的普及使用提供了可能性。

4. 功能的可扩充性

多媒体创作工具强大的多媒体信息处理功能为创作人员提供了完善的开发环境。但是对于某些特殊的需求，如数据库管理功能，如果现有的功能不能实现，这就需要利用多媒体创作工具的扩充功能。一般的多媒体创作工具都提供了其他高级程序设计语言（如 C 语言、SQL 语言等）的接口，创作人员可以利用高级程序设计语言实现自己的特殊需要，然后同多媒体创作工具编辑的多媒体应用程序链接即可。另外，大部分多媒体创作工具支持多媒体创作语言，如 ToolBook 的 OpenScript 语言、Macromedia Director 的 Lingo Script 语言等。利用多媒体创作语言，可以实现对任意多媒体信息的控制，这就使得多媒体创作工具的功能大大增强，对多媒体信息的控制也更加灵活。

5. 缩短开发周期

与利用高级程序设计语言作为工具的开发方式相比，用多媒体创作工具开发多媒体应用程序的周期可大大缩短。这是因为多媒体创作工具对于多媒体信息的呈现已提供了完善的服务功能，创作人员无须再去设计呈现这些信息的函数、过程等，可以把全部精力用于设计多媒体信息呈现方式及其交互操作界面的形式等方面。多媒体应用程序开发的最后调试阶段，也不是烦琐的程序调试，而仅是对多媒体素材呈现顺序的进一步调整。

7.2.2 多媒体创作工具的分类

多媒体创作工具可根据不同方式分类。若按创作特点分类，可分为以下几类。

1. 基于脚本（Script-Based）的创作工具

这类创作工具需提供一套脚本（Script）描述语言或描述符号，设计者用这些语句或符号像写程序那样组织、控制各种媒体元素的呈现、播放。为了便于创作，通常将脚本按页（Page）或卡片（Card）进行组织。

这类工具的典型代表是美国 Asymetrix 公司推出的 Multimedia ToolBook，通常的设计方法是用创作工具中提供的脚本编辑器（如卡片编辑器）通过指令或符号来建立脚本，再利用系统提供的预放（Previewer）系统进行播放，不满意再返回（切换）到脚本编辑器重新设计。为减轻设计者记忆描述语言的负担，一些系统把脚本编辑设计成填表或对话模板方式，设计者只需按格式填写。这类开发环境可以使设计者很容易地一面撰写脚本，一面播放以观察制作效果。

使用脚本语言的优点是可在语句命令中提供变量功能，通过变量的算术运算和逻辑运算，使设计的系统有很大弹性。

2. 基于流程图（Flow-Based）的创作工具

在这类创作工具中，多媒体元素和交互作用提示及数据流程控制都在一个流程图（Flowchart）中进行安排，即以流程图为主干构造结构化的框架或过程，如图 7.1 所示。流程图上的流程线（Line）是数据控制流程，流程线上放置着不同类型的图标（ICON）。图标扮演着类似脚本指令的角色，打开每个图标，就是一个对话框，要求使用者输入内容。在流程线上可对任意图标进行独立编辑和测试。这种工具也称为基于图标的事件驱动工具。流程图方式创作正好符合人的认知规律，可形象地表达大脑中信息加工的过程。

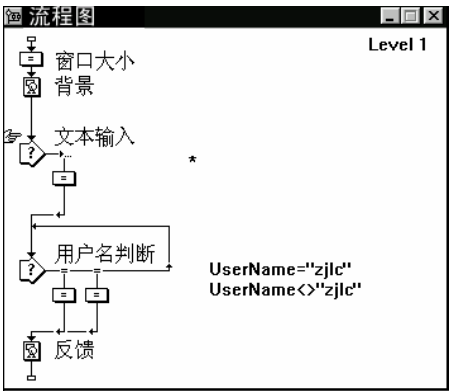


图 7.1 创作流程图

基于流程图的创作工具简化了项目的组织，并使整个设计框架通过流程图一目了然，因此这种编辑方式被称为 **Visual Authoring**，即可视化创作。而且流程图同时可在复杂的系统中作为导航手段，十分有用。这类工具也具有类似脚本指令的优点，可以制作出灵活多变的多媒体节目。**Macromedia** 公司推出的 **Authorware** 是这类工具的典型代表，是目前被公认为交互功能最强的创作工具。它比较成功的应用领域是计算机辅助教学和训练。**Authorware** 采用面向对象的创作，提供直观的图标界面，利用十多种功能图标逻辑结构的布局，体现程序运行的结构，并配以函数和变量完成数据操作，从而取代了复杂的编程语言。

流程图确保了加工过程的确定性，消除了随意性，能确保按照流程图所规定的程序成功地解决问题。但流程图的缺点恰恰又在于其“确定性”。确定性能保证常规性问题的顺利解决，却不能保证创造性地用巧妙的新方法去解决问题。

3. 基于时间序列（Timeline-Based）的创作工具

以时间序列为基础的创作工具是最常见的多媒体编辑软件。主要用来制作电影、卡通片等影视节目，即以看得见的时间线（**Timeline**）来决定事件的顺序和对象演示的时段。这种创作过程除按时间序列安排节目的内容和流程外，还要进行各种媒体资料的同步控制，因此时间序列中可以包括多行道或多频道，以便安排多种对象同时呈现。

这类创作工具适用于信息从头到尾顺序播放的影视应用系统创作。组织的图形帧按预定速度播放，其他媒体元素（如音频、动画等）在时间序列中的给定时间和位置被激活。这类工具的典型代表是 **Macromedia** 公司的 **Director**。

虽然按时间序列在控制媒体的同步上有其独到之处，但对于交互式的操作及在逻辑判断处理上都不如脚本描述和流程图方式那样直观，比较适合于制作交互性不强的商业广告及演示类的节目。

7.3 常用多媒体创作工具介绍

目前多媒体创作工具非常多，常用的多媒体创作工具软件包括 Macromedia 公司的 Director, Authorware 等。Director 和 Authorware 都可以用来制作大型的动画，但是，这两个工具的做法却截然不同，Director 是从导演的角度来安排角色出场的，使用的术语与电影术语非常类似，而 Authorware 的创作是基于流程图的，对角色的安排大多是在流程图中完成的。

1. Director

Director 和 Flash 一样用“Movie”（影片）来比喻整个程序，所有开发程序的过程相当于安排演员在舞台上进行表演的过程，“Director”（英文中是导演的意思）的名字也由此产生。它的基本概念是电影中的“帧”（Frame），制作的最终结果通过舞台显示，而播放的次序和指令通过一个叫“Score”（总谱）的面板进行操作，其他的面板主要为总谱面板提供素材。Director MX 2004 的界面如图 7.2 所示。

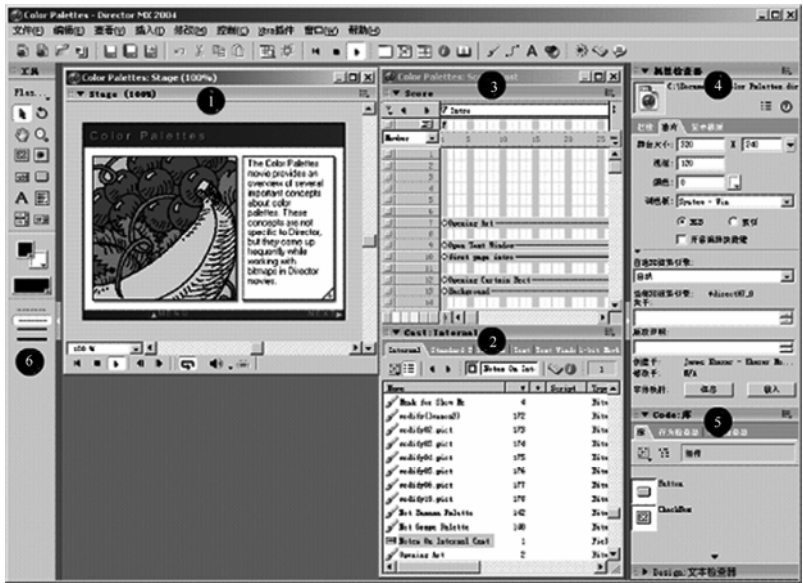


图 7.2 Director MX 2004 界面

（①舞台②演员表③总谱④属性面板⑤库面板⑥工具面板）

2. Authorware

Authorware 是美国 Marcromedia 公司推出的优秀的多媒体制作工具，已被全球大多数多媒体开发厂家采用。Authorware 是一种基于流程图的创作工具，其基本创作思路是：

- ① 新建文件，建立流程图，设置新文件运行环境；
- ② 根据制作脚本要求，选择图标填充于流程线上；

- ③ 组合图标，程序模块化分层，形成整个项目的流程线；
- ④ 打开图标，编辑多媒体对象；
- ⑤ 运行和分段调试程序；
- ⑥ 保存、编辑打包生成可脱离开发环境的可执行文件；
- ⑦ 制作应用项目安装盘或生成 Web 发布文件；
- ⑧ 录制光盘或将软件存储于 Web 服务器中。

用 Authorware 6.5 制作的拼图游戏界面如图 7.3 所示（详见实验 5）。

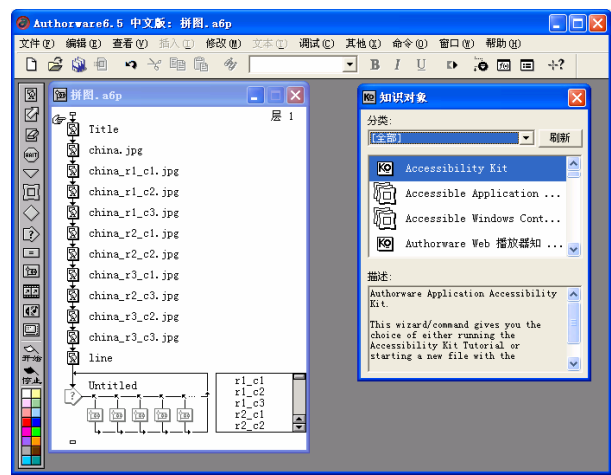


图 7.3 Authorware 6.5 界面

尽管 Director 和 Authorware 是目前公认的最强大的多媒体制作软件，但是，作为一个软件来说，不可能面面俱到，总存在一些不足。为了给用户更多的使用灵活性，Authorware 和 Director 一样，提供了 Xtras 的技术来扩展它的功能。Xtras 的定义是按照 Macromedia 的开发规范 Xtra Development Kit（XDK）开发出来的，专为 Macromedia 的产品提供扩展功能的软件。Macromedia 本身只提供了少量的 Xtras，大多数的 Xtras 都是由第三方开发的。

习 题 7

简答题

- (1) 什么是电子出版物？有什么特点？
- (2) 什么是多媒体创作工具？它具有哪些特点？
- (3) 多媒体创作工具按创作特点分类，可分为哪几类？
- (4) 常见的多媒体创作工具有哪些？
- (5) 采用多媒体创作工具开发多媒体电子出版物与采用软件编程方式实现相比具有什么优势？
- (6) 简述多媒体电子出版物的开发与制作流程。

第 8 章 数字流媒体

随着网络时代的深入，网络上传递的信息种类越来越多，从最初的文字信息发展到目前的文字、图像、声音、视频、动画等几乎所有的信息种类。传递的信息种类的增多，特别是有时需要同时传递多种信息，对计算机网络的数据传输技术也提出了新的要求，在不断提升网络带宽的同时，网络多媒体技术也在不断的发展，包括数据的压缩编码和用于发布媒体的服务器技术。流媒体正是近年来出现的比较新颖、实用的网络多媒体技术。

8.1 流媒体基本概念

8.1.1 流媒体技术概述

在流媒体出现之前，人们若想从网络上观看影片或收听音乐，必须先将影音文件下载至计算机并储存后，才可以播放，不但浪费时间、浪费硬盘空间，而且无法满足消费者的实时需要。

流媒体（Streaming Media）的发展，克服了这些不足。流媒体一般是指把连续的影像和声音信息经过压缩处理后放在网站服务器上，让用户一边下载，一边观看、收听，而不需要等整个压缩文件下载到自己机器后才可以观看的视频/音频传输、压缩技术。

流媒体技术起源于窄带互联网时期，由于经济发展的需要，人们迫切渴求一种网络技术，以便进行远程信息沟通。从 1994 年一家叫做 Progressive Networks 的美国公司成立之初，流媒体便开始正式在互联网上登场亮相。1995 年，他们推出了 C/S 架构的音频服务系统 Real Audio，并在随后的几年内引领了网络流式技术的汹涌潮流。1997 年 9 月，该公司更名为 RealNetworks，相继发布了多款应用非常广泛的流媒体播放器（RealPlayer 系列），在其鼎盛时期，曾一度占据该领域超过 85% 的市场份额。RealNetworks 公司可以称得上是流媒体真正意义上的始祖。但是随后，微软和苹果等都看到了流媒体的大好前景，其强大的竞争攻势一方面令 RealNetworks 感到危机的存在，另一方面也无形中促进了流媒体的迅速发展。

1. 流媒体的特点

（1）启动延时大幅度缩短

用户不用等到所有内容都下载到硬盘上才开始浏览，一部影片在一分钟之内就会显示在客户端上，而且在播放过程中一般不会出现断续的情况。另外，全屏播放对播放速度几乎无影响，但在快进、快退时需要等待。

（2）对系统缓存容量的需求降低

由于 Internet 是以包传输为基础来进行断续的异步传输的，数据被分解为许多包进行传输，动态变化的网络使各个包可能选择不同的路由，故到达用户计算机的时间延迟也就不同。因此，在客户端需要系统缓存来弥补延迟和抖动的影响并保证数据包传输顺序的正确，从而使媒体数据能连续输出（不会因网络暂时拥堵而使播放出现停顿）。虽然流式传输仍需要缓存，

但由于不需要把所有的动画、视频、音频内容都下载到缓存中，因此，对缓存的要求降低。高速缓存使用环形链表结构来存储数据，通过丢弃已经播放的内容，流可以重新利用空出的高速缓存空间来缓存后续尚未播放的内容。

（3）采用特定的实时传输协议

采用 RTSP 等实时传输协议，更加适合动画、视频、音频在网上实时传播。

（4）对时间的敏感性

流媒体的一个重要特征是对时间的敏感性，这正是实时性要求高的应用所必须的，所以这类应用与流媒体密不可分是十分自然的事情。流媒体的实现主要取决于网络带宽和压缩算法的提高。今天，随着网络协议的改善、网络基础设施和压缩技术的发展，流媒体的实现已经变得越来越容易了。

2. 流媒体的传输

（1）顺序流式传输

顺序流式传输就是顺序下载，在下载文件的同时用户可观看在线媒体，在给定时刻，用户只能观看已下载的那部分，而不能跳到还未下载的部分，顺序流式传输不像实时流式传输那样，可在传输期间根据用户连接的速度做调整。由于标准的 HTTP 服务器可发送这种流式传输所需的文件且不需要其他特殊协议，因此，顺序流式传输通常被称做 HTTP 流式传输。顺序流式传输比较适合高质量的短片段，如片头、片尾和广告，由于该文件在播放前观看的部分是无损下载的，因此这种方法保证了电影播放的最终质量。这意味着用户在观看前，必须经历延迟，对较慢的连接尤其如此。

对通过调制解调器发布短片段，顺序流式传输显得很实用，它允许用比调制解调器更高的数据速率创建视频片段。尽管有延迟，但毕竟可发布较高质量的视频片段。

顺序流式文件是放在标准 HTTP 或 FTP 服务器上的，易于管理，基本上与防火墙无关。顺序流式传输不适合长片段和有随机访问要求的视频，如讲座、演说与演示。它也不支持现场广播，严格来说，它是一种点播技术。

（2）实时流式传输

实时流式传输保证媒体信号带宽与网络连接匹配，以使媒体可被实时观看到。实时流式传输与 HTTP 流式传输不同，它需要专用的流媒体服务器与传输协议。

实时流式传输是实时传送，特别适合现场事件，支持随机访问，用户可快进或后退以观看前面或后面的内容。理论上，实时流一经播放就可以不停止，但实际上，有可能发生周期暂停。

实时流式传输必须匹配连接带宽，这意味着在以调制解调器速度连接时，图像质量会较差。而且，由于出错丢失的信息被忽略掉，当网络拥挤或出现问题时，视频质量会很差。如欲保证视频质量，顺序流式传输也许更好。实时流式传输需要特定服务器，如 QuickTime Streaming Server, RealServer 与 Windows Media Server。这些服务器允许你对媒体发送进行更多级别的控制，因而系统设置、管理比标准 HTTP 服务器更复杂。实时流式传输还需要特殊网络协议，如 RTSP (Realtime Streaming Protocol) 或 MMS (Microsoft Media Server)。这些协议在有防火墙时有时会出现问题，导致用户不能看到一些地点的实时内容。

3. 支持流媒体传输的网络协议

支持流媒体传输的网络协议如下。

(1) 实时传输协议 (RTP) 与实时传输控制协议 (RTCP)

RTP (Real-time Transport Protocol) 是用于 Internet 上针对多媒体数据流的一种传输协议。RTP 被定义为在一对一或一对多的传输情况下工作, 其目的是提供时间信息和实现流同步。RTP 通常使用 UDP 来传输数据, 并可以在 TCP 或 ATM 等其他协议上工作。当应用程序开始一个 RTP 会话时, 将使用两个端口: 一个给 RTP, 另一个给 RTCP (Real-time Transport Control Protocol)。RTP 本身并不能为按顺序传送数据包提供可靠的传送机制, 也不提供流量控制或拥塞控制, 它是依靠 RTCP 来提供这些服务的。通常, RTP 算法并不作为一个独立的网络层来实现, 而是作为应用程序代码的一部分。

在 RTP 会话期间, 各参与者周期性地传送 RTCP 包。RTCP 包中含有已发送的数据包的数量、丢失的数据包的数量等统计资料, 因此, 服务器可以利用这些信息动态地改变传输速率, 甚至改变有效载荷类型。RTP 和 RTCP 配合使用, 能以有效的反馈和最小的开销使传输速率最佳化, 因而特别适合传送网上的实时数据。

(2) 实时流协议 (RTSP)

实时流协议 RTSP (Real-time Streaming Protocol) 是由 RealNetworks 和 Netscape 共同提出的, 该协议定义了一对多应用程序如何有效地通过 IP 网络传送多媒体数据。RTSP 在体系结构上位于 RTP 和 RTCP 之上, 它使用 TCP 或 RTP 完成数据传输。HTTP 与 RTSP 相比, HTTP 传送 HTML, 而 RTSP 传送的是多媒体数据。HTTP 请求由客户机发出, 服务器做出响应; 当使用 RTSP 时, 客户机和服务器都可以发出请求, 即 RTSP 可以是双向的。

(3) 资源预留协议 (RSVP)

由于音频和视频数据流比传统数据对网络的延时更敏感, 要在网络中传输高质量的音频、视频信息, 除带宽要求之外, 还需要其他更多的条件。RSVP (Resource Reservation Protocol) 是适用于 Internet 上并满足上述要求的资源预订协议。该协议能预留一部分网络资源 (即带宽), 从而能在一定程度上为流媒体的传输提供 QoS (服务质量)。RSVP 协议位于 OSI 协议模型的传输层, 它是一种网络控制协议, 而非路由协议。它仅沿着数据流所选定的路由来预留资源。RSVP 与路由协议协同工作, 确保端到端之间的传输带宽, 尽量减少实时多媒体通信中的传输延时和抖动。

(4) 多用途因特网邮件扩展协议 (MIME)

MIME (Multipurpose Internet Mail Extension) 不仅用于电子邮件, 还能用来标记在 Internet 上传输的任何文件类型。

Web 服务器和浏览器都是基于 HTTP 协议的, 而 HTTP 都内建有 MIME。HTTP 正是通过 MIME 来标记 Web 上众多的多媒体文件格式的。Web 服务器和浏览器就是用 MIME 来识别流媒体并进行相应处理的。

4. 流媒体系统的组成

流媒体系统主要由以下 5 个部分组成。

① 编码工具: 用于创建、捕捉和编辑多媒体数据, 从而形成流媒体格式。

- ② 流媒体数据：包括视频、音频等多媒体信息。
- ③ 服务器：用于存放和控制流媒体数据。
- ④ 传输网络：适合多媒体传输协议甚至是实时传输协议的网络。
- ⑤ 播放器：供客户端浏览流媒体文件。

5. 流式传输的基本原理

流式传输的实现需要合适的传输协议。由于 TCP 需要较多的开销，故不太适合传输实时数据。在流式传输的实现方案中，一般采用 HTTP/TCP 来传输控制信息，而用 RTP/UDP 来传输实时声音数据。

流式传输的过程一般是这样的：用户选择某一流媒体服务后，Web 浏览器与 Web 服务器之间使用 HTTP/TCP 交换控制信息，以便把需要传输的实时数据从原始信息中检索出来；然后客户机上的 Web 浏览器启动 A/V Helper 程序，使用 HTTP 从 Web 服务器检索出相关参数对 Helper 程序进行初始化。这些参数可能包括目录信息、A/V 数据的编码类型或与 A/V 检索相关的服务器地址。

A/V Helper 程序及 A/V 服务器运行实时流协议（RTSP），以交换 A/V 传输所需的控制信息。与 CD 播放机或录像机所提供的功能相似，RTSP 提供了操纵播放、快进、快退、暂停及录制等命令的方法。A/V 服务器使用 RTP/UDP 协议将 A/V 数据传输给 A/V 客户程序（一般可认为客户程序等同于 Helper 程序），一旦 A/V 数据抵达客户端，A/V 客户程序即可播放输出。实现流式传输一般都需要专用服务器和播放器，其基本原理如图 8.1 所示。

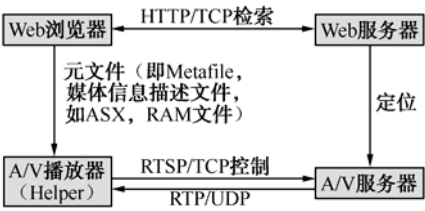


图 8.1 流式传输的基本原理

6. 流媒体播放方式

按照播放模式可分为“点播”和“广播”，按照通信方式可分为“单播”和“组播”。

(1) 点播（On-demand Streaming）

点播连接是客户端与服务器之间的主动连接。在点播连接中，用户通过选择内容项目来初始化客户端连接。用户可以开始、停止、后退、快进或暂停流。点播连接提供了对流的最大控制，但这种方式由于每个客户端各自连接服务器，因此会迅速用完网络带宽。

(2) 广播（Live Streaming）

当以广播模式播放时，客户端被动地接收流。在广播过程中，客户端只能接收流，而不能像在点播中那样暂停、快进或后退来控制流。使用广播模式，在网络上只传输数据包的一个备份，网络上的所有用户都会收到数据包，如图 8.2 所示。

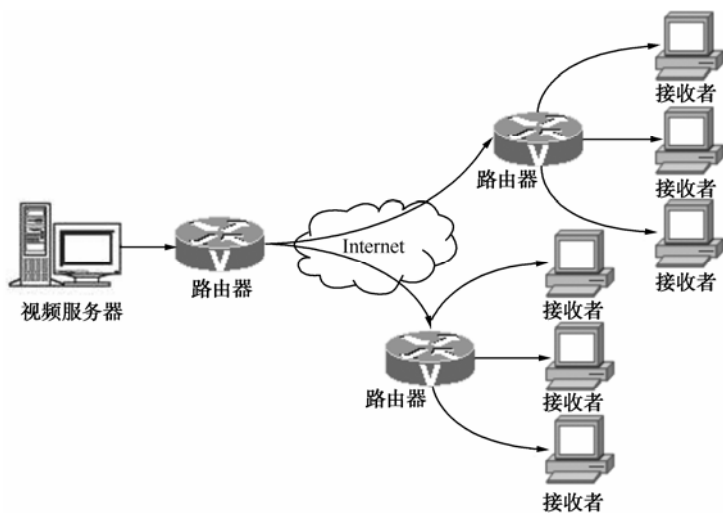


图 8.2 广播

(3) 单播 (Unicast)

在客户端与媒体服务器之间需要建立一个单独的数据通道，从一台服务器送出的每个数据包只能传送给一个客户机，这种传送方式称为单播，如图 8.3 所示。每个用户必须分别对媒体服务器发送单独的查询，而媒体服务器必须向每个用户发送所申请的数据包副本。这种巨大冗余会造成服务器沉重的负担，响应需要很长时间，甚至停止播放，管理人员也被迫购买硬件和带宽来保证一定的服务质量。

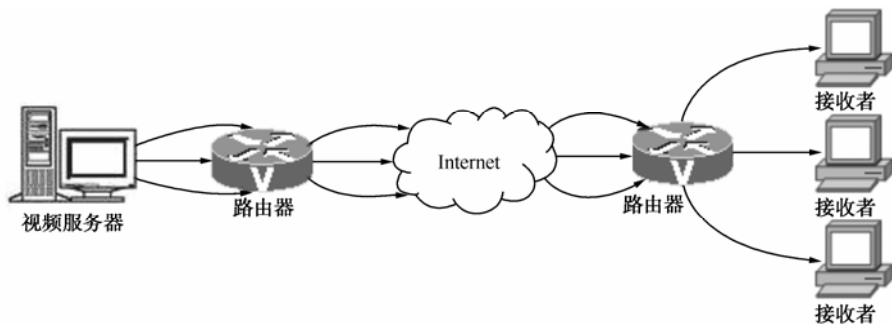


图 8.3 单播

(4) 组播 (Multicast)

IP 组播技术可构建一种具有组播能力的网络，允许路由器一次将数据包复制到多个通道上，如图 8.4 所示。采用组播方式，单台服务器能够对几十万台客户机同时发送连续数据流而无延时。媒体服务器只需要发送一个信息包，而不是多个，所有发出请求的客户端共享同一信息包。信息可以发送到任意地址的客户机，减少了网络上传输的信息包的总量。网络利用效率大大提高，成本大大下降。

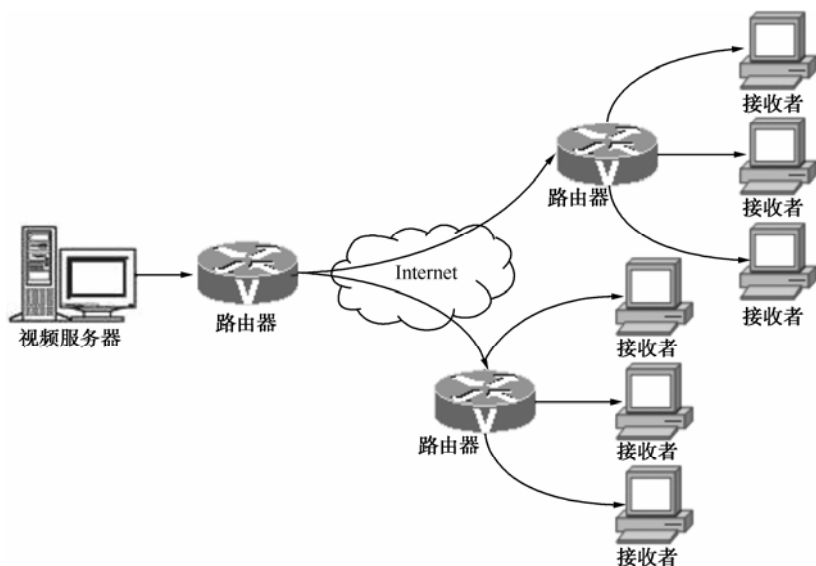


图 8.4 组播

8.1.2 流媒体文件格式

流媒体文件格式是支持采用流式传输及播放的媒体格式。

1. 压缩媒体文件格式

由于视频、音频媒体经数字化后，数据量很大，形成的媒体文件要占用较大的存储空间，不能直接用于传送，因此必须进行压缩编码，使数据量特别大的音、视频等文件显著减小。经过压缩编码后形成的媒体文件，称为压缩媒体文件，压缩媒体文件所采用的格式称为压缩媒体文件格式。压缩媒体文件格式通过压缩编码去掉了大量的冗余信息，但尽量保留了或全部保留了原始媒体的信息。由于压缩过程是自动进行的，并内嵌在媒体文件格式中，通常我们在存储文件时没有注意到这点。该过程如图 8.5 所示。表 8.1 列举了一些常用的视频和音频文件类型。

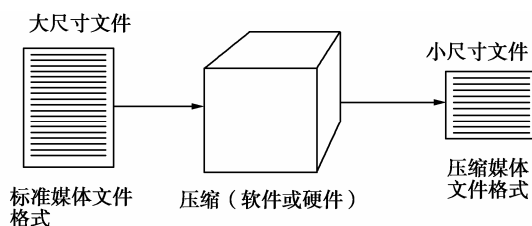


图 8.5 压缩媒体文件的压缩过程

表 8.1 常用视频、音频文件类型

文件扩展名	媒体类型与名称	压缩情况
.mov	Quicktime Video V2.0	可以
.mpg	MPEG—1 Video	有
.mp3	MPEG Layer 3 Audio	有
.wav	Wave Audio	没有

续表

文件扩展名	媒体类型与名称	压 缩 情 况
.aif	Audio Interchange Format	没有
.snd	Sound Audio File Format	没有
.au	Audio File Format (Sun OS)	没有
.avi	Audio Video Interleaved V1.0 (Microsoft Windows)	可以

2. 流式文件格式

流式文件格式经过特殊编码后，可使其适合在网络上边下载边播放（而不是等到下载完整个文件才能播放）。在将压缩媒体文件编码成流式文件时，必须添加一些附加信息，如计时、压缩和版权信息。编码过程如图 8.6 所示。表 8.2 列举了常见的流式文件类型。

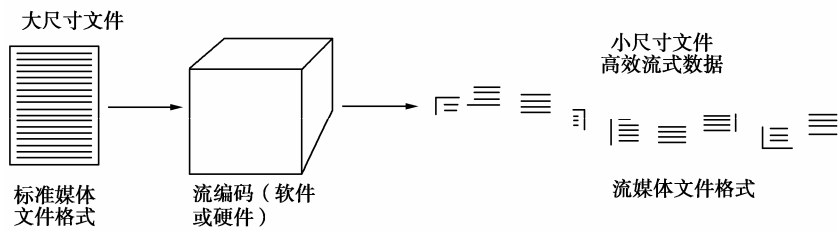


图 8.6 流式文件的编码过程

表 8.2 常见的流式文件格式

文件格式扩展名 (Video/Audio)	媒 体 类 型	公 司 名 称
.asf	Advanced Streaming Format	Microsoft
.wmv	Windows Media Video	Microsoft
.wma	Windows Media Audio	Microsoft
.rm	Real Media	RealNetworks
.rmvb	Real Media Variable Bit Rate	RealNetworks
.ra	Real Audio	RealNetworks
.rp	Real Pix 文件	RealNetworks
.rt	Real Text 文件	RealNetworks
.swf	Flash 文件	Macromedia
.flv	Flash Video 文件	Macromedia
.qt	QuickTime 文件	Apple

在表 8.2 的流式文件格式中，尤其以 Real Video 的 rm、rmvb 格式和 ra 音频格式，Microsoft Media 的 asf、wmv、wma 格式，QuickTime 的 qt 格式及 Flash 的 swf 格式最为常用，现简要介绍如下。

(1) Real Video 的 rm、rmvb 格式和 ra 音频格式

rm 和 ra 格式分别是 RealNetworks 公司开发的一种流式视频 Real Video 和流式音频 Real Audio 文件格式，主要用来在低速率的网络上实时传输活动视频影像。可以根据网络数据传

输速率的不同而采用不同的压缩比率，并在数据传输过程中边下载边播放视频影像，从而实现影像数据的实时传送和播放。客户端是通过 Real Player 播放器进行播放的。

rmvb 格式是由 rm 格式升级延伸而来的，vb 即 vbr，是 Variable Bit Rate（可变比特率）的英文缩写。在播放以往常见的 rm 格式电影时，可以在播放器左下角看到 225kbps 字样，这就是比特率。影片的静止画面和运动画面对压缩采样率的要求是不同的，如果始终保持固定的比特率，会对影片质量造成浪费。而 rmvb 则打破了原先 rm 格式那种平均压缩采样的方式，在保证平均压缩比的基础上，设定了一般为平均采样率两倍的最大的采样率值。将较高的比特率用于复杂的动态画面（如歌舞、飞车、战争等），而在静态画面中则灵活地转为较低的采样率，合理地利用了比特率资源，使 rmvb 在牺牲少部分察觉不到的影片质量情况下最大限度地压缩了影片的大小，最终拥有了近乎完美的接近于 DVD 品质的视听效果。

（2）Microsoft Media 的 asf, wmv, wma 格式

Microsoft Media 的 asf 格式也是一种流行的网上流媒体格式。它是微软为了和现在的 Real Player 竞争而推出的一种视频格式，使用 MPEG—4 的压缩算法，压缩率和图像的质量都很不错。因为 asf 以“流”格式存在，所以它的图像质量比 VCD 差，但比同是视频“流”格式的 rm 格式要好。这种流式文件的使用与 Windows 操作系统是分不开的，使用的播放器是 Microsoft Media Player。

wmv 的全称是 Windows Media Video，也是微软推出的一种采用独立编码方式并且可以直接在网上实时观看视频节目的文件压缩格式，是 asf 格式的升级和延伸。在同等视频质量下，wmv 格式的体积非常小，因此很适合在网上播放和传输。wmv 格式的主要优点包括：本地或网络回放、可扩充的媒体类型、部件下载、可伸缩的媒体类型、流的优先级化、多语言支持、环境独立性、丰富的流间关系及扩展性等。

wma 的全称是 Windows Media Audio，它是微软公司推出的与 mp3 格式齐名的一种新的音频格式。在低比特率时，效果好过 mp3；在高比特率时，wma 作用不大。因压缩率高，wma 文件适宜于网络下载。

wma 及 wmv 的压缩数据都是用 asf 的文件格式存储的，纯音频的 asf 和 wma 一样。

（3）QuickTime 的 qt 格式

QuickTime 的 qt 格式是 Apple 公司开发的一种音频、视频文件格式。该格式具有先进的音频和视频功能，有包括 Apple Mac OS，Microsoft Windows 95/98/NT/2000，Linux 在内的所有主流计算机操作系统的支持。QuickTime 文件格式支持领先的集成压缩技术，提供 150 多种视频效果，并配有提供了 200 多种 MIDI 兼容音响和设备的声音装置。能够通过 Internet 提供实时的数字化信息流、工作流与文件回放功能，此外，QuickTime 还采用了一种称为 QuickTime VR（简称 QTVR）技术的虚拟现实（Virtual Reality，VR）技术。QuickTime 以其领先的多媒体技术和跨平台特性、较小的存储空间要求、技术细节的独立性及系统的高度开放性，得到业界的广泛认可，目前已成为数字媒体软件技术领域的事实上的工业标准。

（4）Flash 的 swf 格式

Flash 的 swf 格式是 Macromedia 公司的流式动画格式。由于其体积小，功能强、交互能力好等特点，越来越多地应用到网络动画中，流行于 Internet 上，在客户端安装 Flash Player 插件即可播放。

（5）Flash 的 flv 格式

随着 Flash MX 的推出，Macromedia 公司开发了属于自己流式视频格式——flv。flv 流媒

体格式是一种新的视频格式，全称为 Flash Video。由于它形成的文件极小、加载速度极快，使得网络观看视频文件成为可能，它的出现有效地解决了视频文件导入 Flash 后，使导出的 swf 文件体积庞大，不能在网络上很好的使用等缺点。

自从 Youtube (Youtube 一直以来在使用 flv 流媒体格式) 火了以后，flv 格式也开始火爆，目前国内视频分享网站，比如土豆、56、优酷等都使用 flv 这个文件技术来实现的。

8.1.3 流媒体技术的应用

互联网的迅猛发展和普及为流媒体业务发展提供了强大的市场动力，流媒体业务正变得日益流行。流媒体技术广泛用于新闻出版、证券、娱乐、电子商务、远程培训、视频会议、远程教育、远程医疗等互联网信息服务的方方面面，它的应用将为网络信息交流带来革命性的变化，流媒体技术改变了传统互联网的呆板形象，丰富了互联网的功能，成为一种有强大吸引力的新媒体。

1. 流媒体技术的主要应用

(1) 视频点播

随着计算机技术的发展，流媒体技术越来越广泛地应用于视频点播 VOD 系统。现在，很多大型的新闻娱乐媒体，如中央电视台和一些地方电视台等，都在互联网上提供基于流媒体技术的节目。

目前 VOD 技术逐渐趋于完善，VOD 技术广泛应用于局域网及有线电视网。流媒体的视频直播应用突破了网络带宽的限制，实现了在低带宽环境下的高质量影音传输，其中的智能流技术保证不同连接速率下的用户，可以得到不同质量的影音效果。但音、视频文件的大容量仍然阻碍了 VOD 技术的进一步发展。由于服务器端不仅需要大容量的存储系统，同时还要承担大量数据的传输，因而服务器根本无法支持大规模的点播。同时，由于局域网中的视频点播覆盖范围小，用户也无法通过 Internet 等网络媒介收听或观看局域网中的节目。

(2) 视频会议

视频会议是流媒体技术的一个商业用途，采用流媒体格式传送音、视频文件，使用者不必等待整个影片传送完毕就可以实时、连续地观看，虽然在画面质量上有一些损失，但就一般的视频会议来讲，并不需要很高的图像质量。当然，流媒体技术并不是视频会议的必须选择，但为视频会议的发展起了重要的推动作用。

通过流媒体进行点对点的通信，最常见的就是可视电话。只要两端都有一台接入 Internet 的计算机和一个摄像头，在世界任何地点都可以进行音、视频通信。此外，大型企业可以利用基于流媒体的视频会议系统来组织跨地区的会议和讨论。

(3) 远程教育

计算机的普及、多媒体技术的发展及 Internet 的迅速崛起，给远程教育带来了新的机遇。越来越多的远程教育网站开始采用流媒体作为主要的网络教学方式。在远程教学过程中，最基本的要求就是将信息从教师端传到远程的学生端，需要传送的信息可能是多元的，如视频、音频、文本、图片等。将这些信息从一端传送到另一端是实现远程教学需要解决的问题，在当前网络带宽的限制下，流式传输将是最佳选择。学生在家通过一台计算机、一条电话线、一个调制解调器就可以参加远程教学。教师也无须另外做准备，授课的方法基本上与传统授课方法相同，只不过面对的是摄像头和计算机而已。

使用流媒体的 VOD 技术还可以进行交互式教学，达到因材施教的目的。如 Real System，

Flash, Shockwave 等技术就经常应用到网络教学中,学生可以通过网络共享学习经验。大型企业可以利用基于流媒体技术的远程教育对员工进行培训。

(4) Internet 直播

随着宽带网的不断普及和流媒体技术的不断发展,冲浪者能够在 Internet 上直接收看体育赛事、商贸展览等,厂商可以借助网上直播形式将自己的产品和活动传遍全世界。网络带宽问题的改善促进了 Internet 直播的发展,Internet 直播已经从实验阶段走向实用,并能够提供较满意的音、视频效果。

流媒体技术的发展,实现了在低带宽环境下提供高质量的音、视频信息;保证不同连接速率下的用户能够得到不同质量的音、视频效果;可减少服务器端的负荷,同时最大限度地节省带宽,在 Internet 直播中充当着重要角色。

(5) 校园视频网

校园网的建设近几年来也逐渐呈现出蓬勃向上的态势,随着多媒体技术的不断发展,特别是多媒体传输技术的突破,使网络多媒体教学得以实现。现在已经有成熟的产品用来组建校园视频网,提供实时广播、定时广播、视频点播 3 种通信模式。

(6) 基于 P2P (Peer To Peer) 的流媒体应用

宽带技术的普及使得人们在互联网上获得了更多的娱乐,比如 VOD 在线点播、影视音乐的下载,可以说人们不再像以前那样只限于浏览简单的网页,收发电子邮件等。但随着网民数量的不断增加,影视资源的获取,不管是在线观看还是下载,都受到了服务器带宽的严重制约,要么就是看得断断续续的,要么就是下半天也下载不完。

近年来,随着 eMule, Bittorrent 等 P2P 应用软件的出现,对 P2P 网络的研究已广泛展开。在 P2P 网络中,每个终端同时具有客户端和服务端功能,每个节点将接收的内容缓存并提供给其他请求节点,这充分挖掘了 Internet 上被忽视的客户机资源,在利用率、扩展性、容错等方面具有巨大的潜力。

P2P 技术对于恶劣的网络环境起到了很好的适应作用,基于 P2P 技术进行流媒体传输应运而生,通过 P2P 技术来进行流媒体播放,打破了传统的 C/S (客户机/服务器) 模式,使服务分散化,从而减轻服务器负载,可以让一个播放节点同时和数个播放节点交换 (提供或索取) 数据,使得带宽的占用达到最大化,获得最佳的播放效果。Mysee、CoolStreaming、PPStream、PPLive、腾讯 QQLive 及 MOP 的 Pcast 这些都是利用 P2P 做网络流媒体点播与直播的典范。

2. 流媒体技术的发展及影响

流媒体技术的发展依赖于网络的传输条件、媒体文件的传输控制、媒体文件的编码压缩效率及客户端的解码等几个重要因素。其中任何一个因素都会影响流媒体技术的发展和运用。不同于 IP 网络的其他新业务,流媒体是网络和数字媒体技术的整合,能够最大限度地发挥网络和数字媒体各自的长处,但也不可避免地继承了网络和数字媒体的缺陷,例如 IP 网络领域服务质量保证的缺乏、数字媒体领域 DVD 版权保护的失败等。

流媒体增值业务平台的构筑将使其应用更加广泛,潜在的客户群体包括电信、广电、智能小区、智能楼宇、校园网、酒店、企业、公安等。应用领域涵盖教育、金融、证券、会议、电子商务、娱乐、监控等。

当前,信息产业界正在经历全球流媒体市场爆炸性的增长。Internet 上已有许多广播电台和电视台实现了网上点播。世界上几大有影响的媒体,如 BBC, CNN, ABC, NBC 等,都

在网上开通了自己的网上广播。随着网络宽带化的飞速发展，作为第 4 媒体（网络媒体）的 Internet 必将超过另外 3 种媒体（报纸、广播、电视）成为主流的信息交换平台。对流媒体业务平台安全性的认识需要上升到国家安全的高度来认识，独立自主开发安全可靠、具有自主知识产权的流媒体增值业务平台已经成为当务之急。

8.2 流媒体处理技术

到目前为止，Internet 上使用较多的流媒体技术主要有 RealNetworks 公司的 Real System、Microsoft 公司的 Windows Media Services 和 Apple 公司的 QuickTime，它们是网上流媒体传输系统的 3 大主流，其中，Real System 和 Windows Media Services 使用最为广泛。

8.2.1 Real System

1. Real System 的系统构成

Real System 由媒体内容制作工具 Real Producer、服务器 Real Server、客户端播放器 Real Player 组成，如图 8.7 所示。Real System 流媒体文件包括 Real Audio，Real Video，Real Presentation 和 Real Flash 4 类文件，分别用于传送不同的文件。

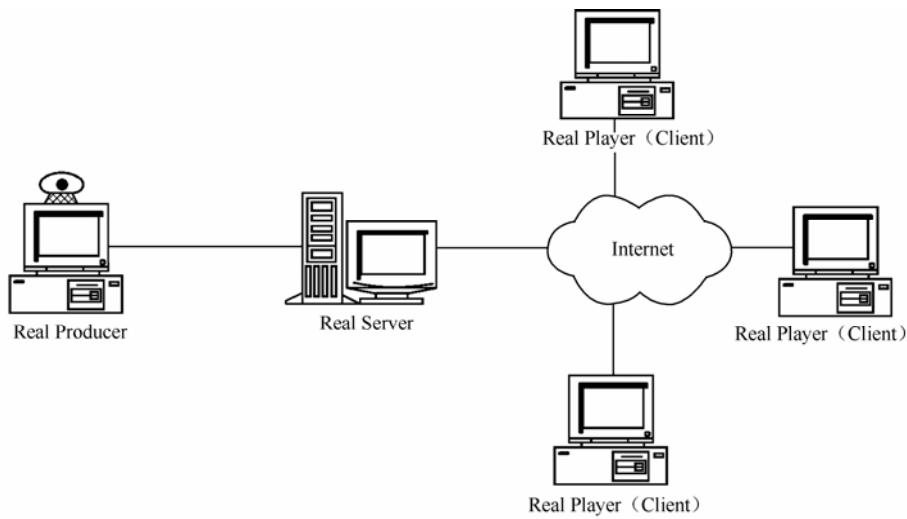


图 8.7 Real System 的系统构成

① Real Producer。Real Producer 主要利用压缩来制作多媒体内容文件。例如，把非 Real 格式的其他音频、视频、动画等多媒体文件转换成 Real Server 可以认识并进行流式广播的 Real 格式；也可以实时压缩现场信号并传给 Real Server 进行现场流式直播；同时，还可以创建 Real 同步多媒体集成语言（Synchronous Media Integration Language, SMIL）文件，并把事先制作的各种剪辑进行合成（如控制每个剪辑的播放时间，对于文字、视频还可以安排它们的显示位置等）。

② Real Server。Real Server 就像 Web 服务器传送网页一样，能够把事先制作好的媒体内容通过因特网传送给客户。它是一种 Internet 和 Intranet 上的流式传输引擎，利用该引擎可以

在客户端实时视听直播节目。并且，Real 公司对外开放自己 Real Server 的内部结构，提供二次开发的接口，允许第三方厂商对 Real Server 进一步开发来满足客户的功能需求。同时，Real Server 还提供了监控客户端收听情况的监控功能模块，如果装上 Real Report 模块，还可以打印出客户端收听情况的详细报告。

③ Real Player。Real Player 是 RealNetworks 公司的 Internet 在线播放器，用来请求并播放 Real Server 传送来的媒体节目。Real Player 不仅能播放 RealNetworks 自身产生的流媒体格式文件（如*.ram, *.rm, *.rmvb 等），而且还能播放众多多媒体格式的文件（如 SMIL, Flash, GIF, QuickTime, MP3 文件等）。

2. 智能流和分流技术

(1) 智能流技术

Real System 采用智能流（SureStream）技术自动并持续地调整数据流的流量，以适应实际应用中的各种不同网络带宽的需求，从而轻松地在网上实现视频、音频和三维动画的回放。智能流技术是 RealNetworks 公司具有代表性的技术，其工作原理如图 8.8 所示。智能流技术从 Real System G2 版本开始引入，它通过 Real Server 将 A/V 文件以流的方式传输，然后利用 SureStream 技术，根据客户端不同的带宽，让传输的 A/V 信息自动适应带宽，并始终以流畅的方式播放。

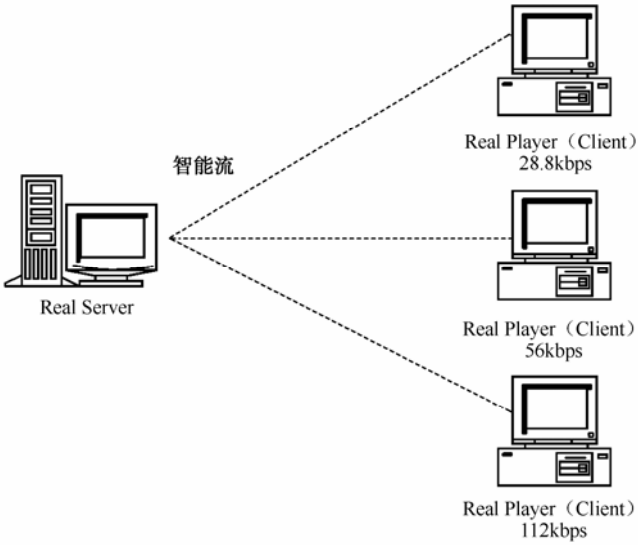


图 8.8 智能流的工作原理图

由于软件、设备、数据传输速度的不同，因而用户可以在不同的带宽下观看节目。当用户请求一段节目内容时，它同时将其带宽能力信息发送给 Real Server。Real System 的编码工具能把媒体记录成不同速度并把它们存储在单一文件中，这样的 Real Audio 和 Real Video 文件叫做 SureStream（智能流）文件。收到用户请求的 Real Server 服务器同时得到用户的带宽信息，然后根据用户带宽确定文件中对应带宽的部分，把与请求对应的最高带宽的部分流传输到用户，这样用户就能获得最高的传输质量。另外，Real Server 还可以根据带宽的变化自动调节流的速率。

如果智能流文件中没有与用户的请求带宽相匹配的编码部分，则 Real Server 就会向用户发送一条匹配带宽的消息，只有使能传输面向不同带宽编码的 SureStream 文件。

(2) 分流 (Splitting) 技术

Real Server 使用分流技术在服务器之间传输直播数据, 如图 8.9 所示。分流方法可以解决 Real Server 超负荷的问题, 使得客户端可以就近访问 Real Server 服务器, 从而获得更好的访问质量并减少带宽使用, 以服务更多用户。Splitting 技术可以采用 UDP 单播、UDP 组播和 TCP 3 种方式进行通信。通过分流, 可使一个或多个 Real Server 服务器加入到发送服务器 (Transmitter) 中, 以分散 Transmitter 的流数量, 而不是所有的请求都到达一个 Real Server 服务器。

在图 8.9 中, 实况内容源处的 Real Server 是发送服务器 (Transmitter), 它将实况播放给其他 Real Server 服务器接收, 负责接收的 Real Server 服务器 (Receiver) 一般更靠近访问者。网页上的链接指向接收的 Real Server 服务器而不是发送服务器, 当用户单击链接时, 接收服务器识别出特定的 URL, 然后把从发送服务器传来的视频流转播给用户。

当 Transmitter 开始播放实况流时, 便将节目广播给所有的 Receiver。当用户从 Receiver 上请求一个播出节目时, Transmitter 和 Receiver 之间就已经建立了一个连接, 此时, 播出节目便会立即发送到用户。

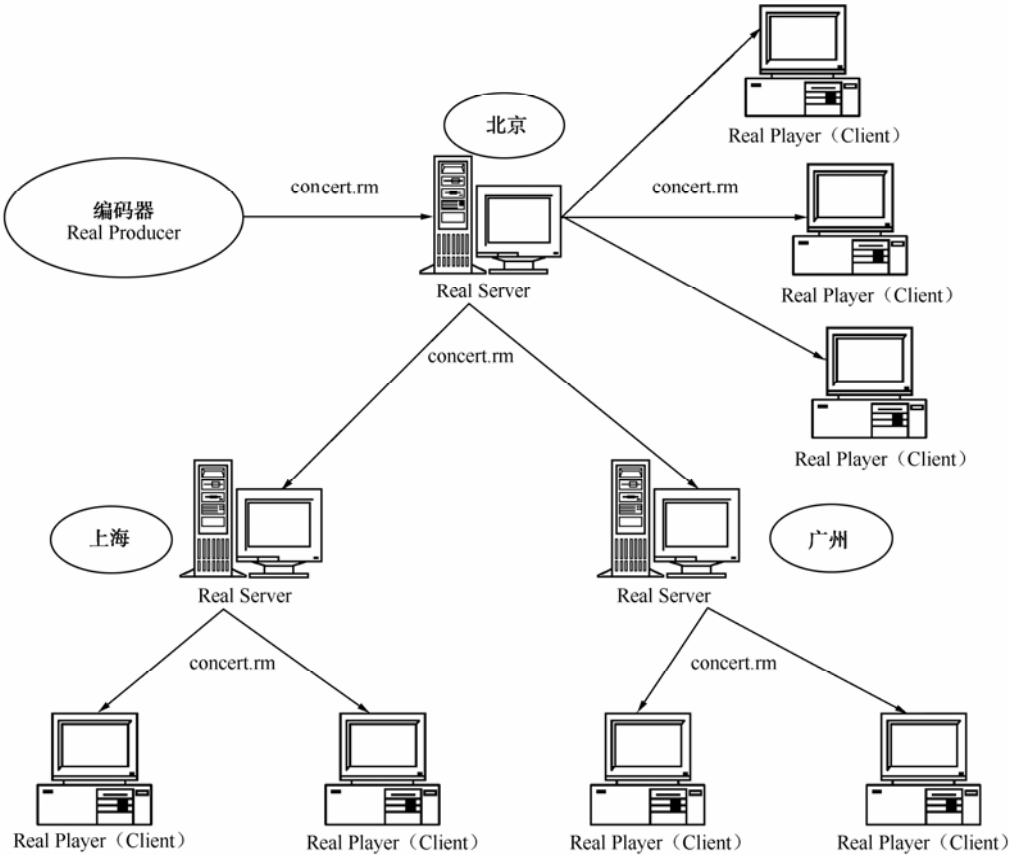


图 8.9 分流技术示意图

8.2.2 Windows Media Services

1. Windows Media Services的系统构成

Windows Media Services 是由微软公司推出的一种能适应多种网络带宽条件的流媒体发

布平台。它包括流媒体的制作、发布、播放和管理等一整套解决方案，如图 8.10 所示，另外，还提供了开发工具包（SDK）供二次开发使用。Windows Media Services 采用基于开发标准的格式 ASF（Advanced Streaming Format，高级流格式），ASF 文件可以由普通视频文件转换而成，特别适合在 IP 网上传输。Windows Media Services 由 Windows Media Tools, Windows Media Server 和 Windows Media Player 工具构成。

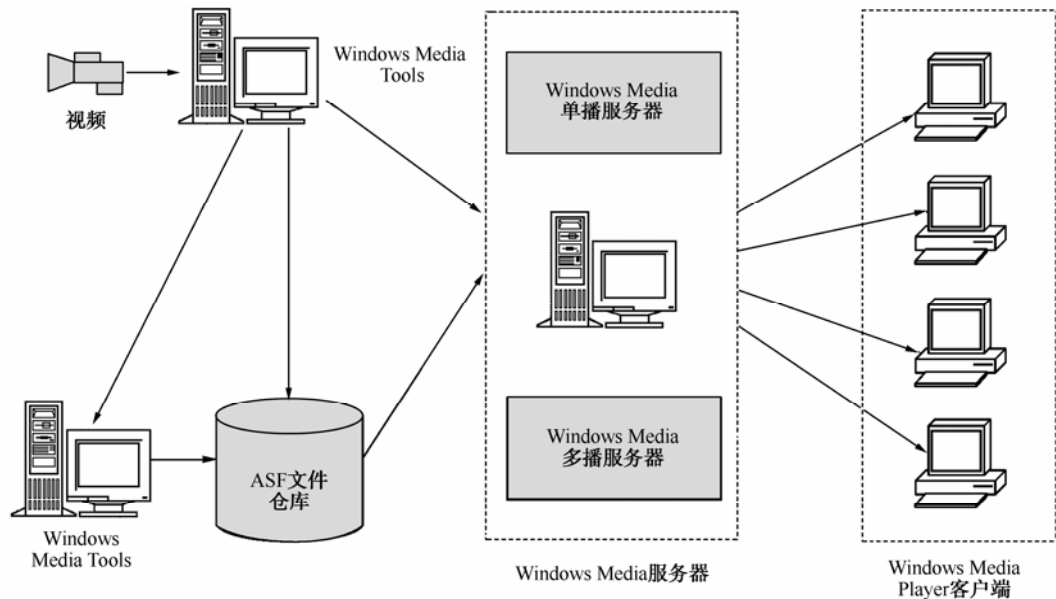


图 8.10 Windows Media Services 的系统构成

① Windows Media Tools: Windows Media Tools 是整个方案的重要组成部分，它提供了一系列的工具来帮助用户生成 ASF 格式的多媒体流（包括实时生成的多媒体流）。这些工具可分成编码工具和编辑工具两大类。

编码工具的主要功能是将各种信号源及其他类型的文件进行压缩编码，并转换为 ASF 格式的多媒体流。如将 AVI 或 MOV 等视频文件转换为 ASF 流式文件；将 Wave 或 MP3 音频文件转换为 ASF 流式文件；利用 Powerpoint 的插件 Presenter 将 Powerpoint 制作的电子幻灯片转换成 ASF 文件。

编辑工具主要用于对 ASF 格式的多媒体流信息进行编辑与管理。其主要功能包括：在 ASF 流中加入流的名称（Title）、描述（Description）等信息（它们会在 ASF 流播放时显示在播放器的信息窗口中）；对 ASF 文件进行编辑管理并设置影片的长短；在 ASF 流中添加标记（Marker），通过标记浏览器可以快速地跳转到指定的位置，这给信息的快速浏览提供了相当便利的手段；在 ASF 流中添加 URL 类型的描述（Script），当浏览器播放到该 ASF 文件的 Script 位置时，播放器会自动地启动 Internet 浏览器并链接到指定的 URL；对 ASF 流进行检查并改正错误等。值得说明的是，在 Marker 和 Script 两个功能的实现过程中，还需要编辑一定的网页。

② Windows Media Server: Windows Media Server 集成在 Windows 2000 Server 中，是 Media Services 的核心，它对外提供 ASF 流式媒体的网络发布服务，主要功能是完成流信息的播放。播放形式包括单播、广播和多播。

Windows Media Server 系统还提供了一套主页形式的管理工具，可以方便地对服务器进行远程管理，完成服务器配置，监控运行时的各种事件、流量及控制客户访问的日志记录等。

③ Windows Media Player: Windows Media Player 是一种通用媒体播放器，可以接收音频、视频和目前较流行的多种混合格式媒体，支持流媒体、在线视听、观看实时新闻等。其支持的媒体格式有 MIDI、MP3、MPEG 视频文件、Microsoft 流式文件、QuickTime 文件等。Windows Media Player 可以独立使用，也可以以 ActiveX Control 的形式嵌入到浏览器或其他应用程序中。

2. Windows Media服务协议

Windows Media 服务协议包括 MMS (Microsoft Media Server, 微软媒体服务器) 协议、MSBD (Media Stream Broadcast Distribution, 媒体流广播发布协议) 协议和 HTTP (Hypertext Transportation Protocol, 超文本传输协议) 协议，这些协议之间的通信过程如图 8.11 所示。

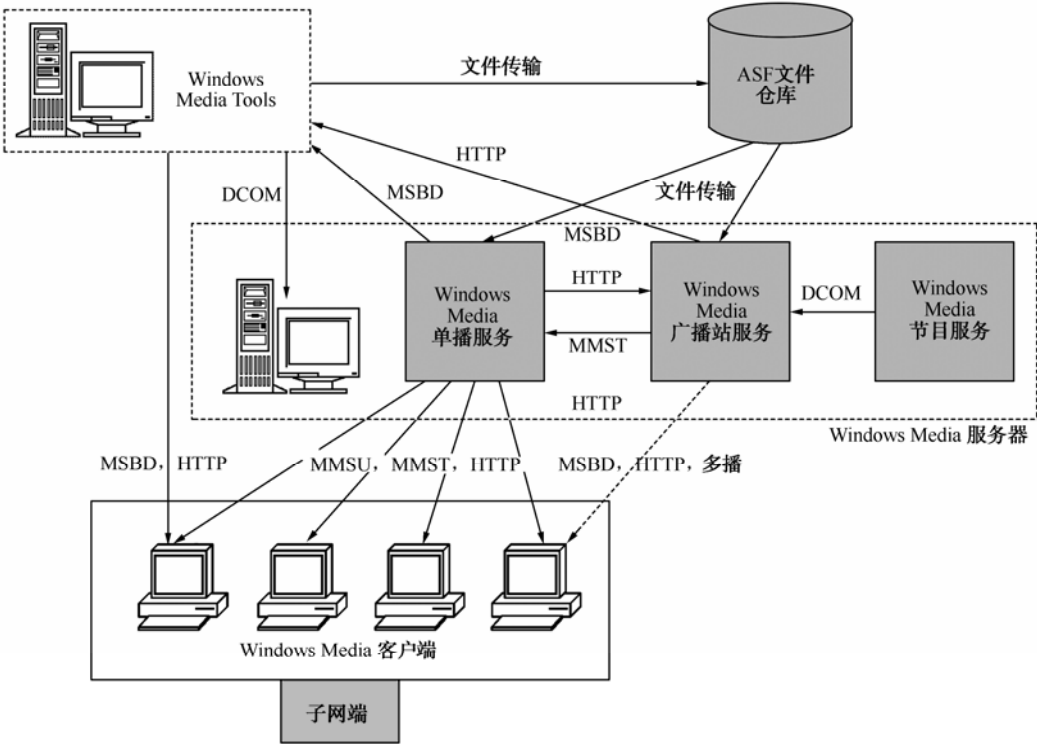


图 8.11 Windows Media 服务协议的通信过程

Windows Media 节目服务 (Windows Media Program Service): 将 Windows Media 流组合至 Windows Media 广播站服务连续节目内。

Windows Media 广播站服务 (Windows Media Station Service): 为传输 Windows Media 内容提供多播和分发服务。

Windows Media 单播服务 (Windows Media Unicast Service): 将 Windows Media 流点播内容提供给网络客户。

(1) MMS 协议

MMS 协议是用来访问并接收 Windows Media 服务器中 ASF 文件的一种协议。MMS 协议用于访问 Windows Media 发布点上的单播内容，是连接 Windows Media 单播服务的默认方法。若观众在 Windows Media Player 中输入一个 URL 以连接内容，而不是通过超级链接访问

内容，则他们必须使用 MMS 协议引用该流。

当使用 MMS 协议连接到发布点时，可使用协议翻转以获得最佳连接。协议翻转始于服务器试图通过 MMSU 连接客户端(MMSU 是 MMS 协议结合 UDP 的数据传送方式)。如果 MMSU 连接不成功，则服务器使用 MMST (MMST 是 MMS 协议结合 TCP 的数据传送方式) 连接。

从独立的 Windows Media Player 连接到发布点，必须指定单播内容的 URL。如果内容在主发布点点播发布，则 URL 由服务器名和 ASF 文件名组成。例如，mms://windows_media_server/sample.asf，其中 windows_media_server 是 Windows Media 服务器名，sample.asf 是点播的 ASF 文件名。如果有实时内容要发布，则该 URL 由服务器名和发布点别名组成。例如，mms://windows_media_server/LiveEvents，这里 windows_media_server 是服务器名，而 LiveEvents 是发布点名。

(2) MSBD 协议

MSBD (Media Stream Broadcast Distribution，媒体流广播发布协议) 是基于 TCP 的协议，主要用于在 Windows Media 编码器和 Windows Media 服务器之间分发流文件，也可以用于在服务器之间分发流文件，而 MMS 用于传输单播数据。MSBD 对于测试客户端、服务器连接和 ASF 内容品质很有用处，但不能作为接收 ASF 内容的主要方法。Windows Media 编码器最多可支持 15 个 MSBD 客户端，而一个 Windows Media 服务器最多可支持 5 个 MSBD 客户端。

(3) HTTP 协议

可以配置 Windows Media 服务器使用 HTTP 协议将内容转化为流。使用 HTTP 流可以帮助克服防火墙障碍，因为大多数防火墙允许 HTTP 通过。HTTP 流可用来由 Windows Media 编码器通过防火墙到 Windows Media 服务器，并可用于连接被防火墙隔离的 Windows Media 服务器。若用同一台计算机既作为 Web 服务器又运行 Windows Media 服务，例如，Microsoft Internet 信息服务 (IIS)，请确保在端口 80 无冲突。

8.2.3 QuickTime

Apple 公司于 1991 年开始发布的 QuickTime 几乎支持所有主流的个人计算机平台和各种格式的静态图像文件、视频和动画格式。可以直接播放 QuickTime 电影、AVI、AIFF、Macromedia Flash 文件，基于 HTTP, RTP, RTSP, FTP 的在线影音、视频。QuickTime 包括服务器 QuickTime Streaming Server (QTSS)、播放器 QuickTime Player、制作工具 QuickTime Pro，如图 8.12 所示。

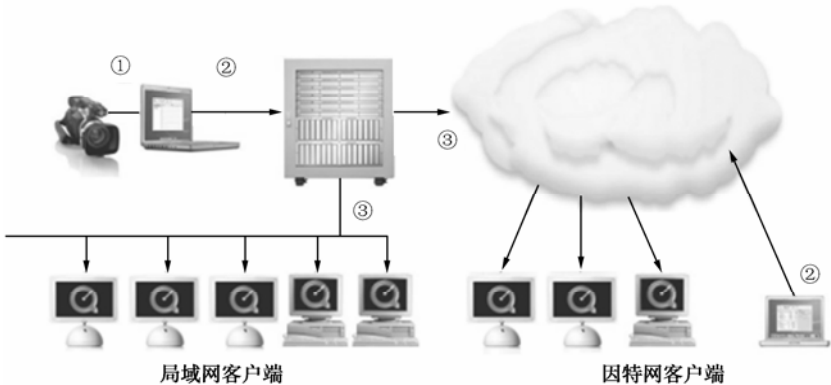


图 8.12 QuickTime 的系统构成

- ① QuickTime Pro 捕获实时内容并进行流式编码。
- ② 将编码好的流文件分发给 QuickTime Streaming Server 并进行节目表的更新。
- ③ QuickTime Streaming Server 通过 RTP/RTSP 协议将流文件传送到本地网络或 Internet 上。

QuickTime 是数字媒体的工业标准，就目前来看，在西方国家中，使用 QuickTime 流媒体技术比较普遍。在中国，RealMedia 和 Windows Media 流媒体技术使用量较大。RealMedia 和 Windows Media 流媒体技术各自提供了一个完全开放的网络视频、音频开发平台，包含了交互式流媒体应用程序的各个方面，为建立强大的端到端的视频、音频流传输和播放提供了解决方案。两者都推出了完善的制作和转换工具，能将各种数据类型从它们原来的格式流化，以流媒体的格式在网上进行传送，完成点播和广播功能。播放质量是流媒体技术的重要方面，但还应考虑到其他几个技术环节，即编码工具、制作工具、播放器和服务器。RealNetworks 和 Microsoft 公司都为用户提供了一整套工具，Windows Media 的工具捆绑在 Windows 操作系统中，有的可以从网上免费下载，不需用户额外购买。在 RealMedia 的工具中，Basic 版本可以从网上免费下载，但 Plus，Pro 版本必须购买，且与用户数有关。表 8.3 显示了 3 种流媒体技术的比较。

表 8.3 3 种流媒体技术的比较

项 目	QuickTime 技术	Real Media 技术	Windows Media 技术
文件格式	MOV	RM, RMVB	ASF, WMV
专用媒体服务器(价格)	贵	贵（与用户数有关）	免费（Windows 2000 Server 自带）
操作系统	Windows 系列及 Mac OS	Windows 系列、Mac OS、Sloaris、Linux	Windows 系列
人-机交互能力	非常好	一般	一般
编码工具费用	价格昂贵，且要多个方面配合	基本编码器免费，高级功能需要购买	编码器全免费

总的来说，3 种流媒体解决方案各有特色，选择时需要综合考虑价格、产品性能、可扩展性和服务等多方面因素。

8.3 流媒体的下载

现在，越来越多的人开始使用宽带，流媒体的在线点播日益风行，许多时候因为网速不稳定，在线点播就会丢帧或者时常缓冲，下载后观看就没有这个问题。但是，流媒体不能用网络蚂蚁这些传统下载工具下载，也不能用浏览器下载。因为这些流媒体文件的连接协议并非常见的 HTTP 或者 FTP，因此根本无法用 NetAnts 等常规的软件下载，而且流媒体遵守的传输协议还没有统一，微软的 MMS 与 RealNetworks 的 RSTP 仍在争夺传输标准。

8.3.1 HiDownload

HiDownload (HD) 是一款可以从 Web, FTP 站点, MMS 和 RTSP 站点下载文件的下载管理工具。允许你以最快的速度下载电影、音乐等文件，同时具备搜索 mp3，分析 MMS/RTSP 下载地址的功能。HiDownload 的界面如图 8.13 所示。

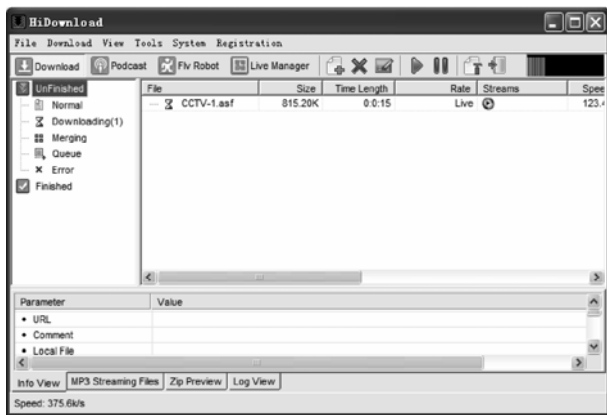


图 8.13 HiDownload 界面


HiDownload 的简单下载流程是：先在浏览器找到想要下载的流节目，在其链接上面单击鼠标右键，选“复制快捷方式”把连接地址复制到剪贴板，然后运行 HiDownload，单击“File->Add”或 （添加按钮），可以看见软件已经自动把你刚才的地址复制到待下载的位置框，再选择文件的保存路径和文件名称，至于下载协议如果不清楚就让 HiDownload 自动判别，如图 8.14 所示。



图 8.14 HiDownload 下载设置

以上步骤完成后单击“OK”按钮，HiDownload 分析地址无误后就会开始下载了。

8.3.2 Net Transport

Net Transport——网络传送带（又名影音传送带），它是国内第一个实现 HTTP，HTTPS，FTP，MMS 和 RTSP 的下载利器，而且支持这些协议的多线程断点续传，这使得它在下载流媒体文件时速度一流，鲜有对手。

Net Transport 的主界面与网际快车非常相似，如图 8.15 所示。

对流媒体文件的链接单击右键选择“使用网络传送带下载”菜单，就会出现如图 8.16 所示的添加新的下载任务窗口。可依据自身的带宽设置合适的可同时下载的块数（线程数），最大可填 10，一般情况下设为 5 就已经足够了（MMS 协议流每秒的字节数差不多是固定的，这就决定了 MMS 流本身不能充分利用用户所有的带宽，所以多线程技术在缩短 MMS 流下载时间方面的效果立竿见影），最后单击“确定”按钮开始下载。

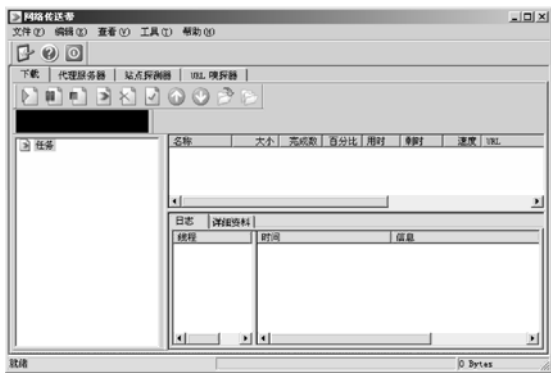


图 8.15 Net Transport 主界面

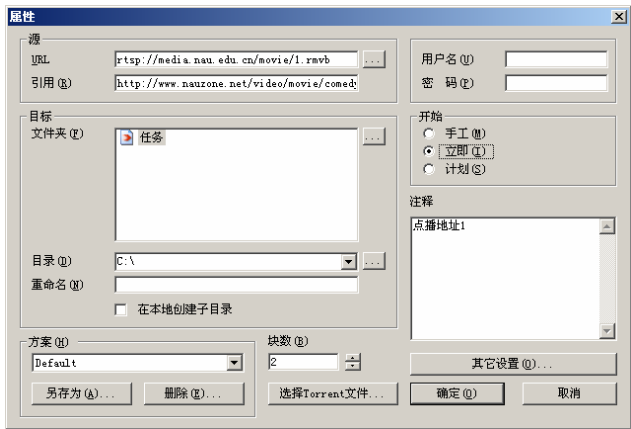


图 8.16 添加新的下载任务窗口

当然，Net Transport 还有其他的功能，比如监视浏览器单击、带有站点探测器等。不难看出，Net Transport 这个功能强大的下载工具，下载流媒体速度一流，尤其在宽带上特别明显。

注意：下载的 ASF、WMV 或者 RM 文件经常会出现问题，可用软件将其修复。常见的修复软件主要有 ASFFix 和 RMFix 等，它们可以对不能播放或未下载完全的流式文件进行修复，还可以对不能拖动播放的流式文件进行重建索引数据，使它能够随意播放。

习 题 8

1. 填空题

- (1) 流媒体播放方式按照播放模式可分为_____和_____；按照通信方式可分为_____和_____。
- (2) _____协议能预留一部分网络资源（即带宽），从而能在一定程度上为流媒体的传输提供 QoS（服务质量）。
- (3) 到目前为止，Internet 上使用较多的流媒体技术主要有_____、_____
_____和_____，它们是网上流媒体传输系统的 3 大主流。

2. 简答题

- (1) 什么是流媒体？具有哪些特点？
- (2) 支持流媒体传输的网络协议有哪些？
- (3) 流媒体系统由哪几部分组成？
- (4) 流式传输的基本原理是什么？
- (5) 流媒体的播放方式有哪些？
- (6) 常见的流媒体文件格式有哪些？
- (7) 流媒体技术的主要应用有哪些？
- (8) 到目前为止，Internet 上使用较多的流媒体技术主要有哪三大主流技术？各有什么特点？

第9章 多媒体光盘存储系统

多媒体的信息包括文本、图形、图像、声音等，由于这些媒体的信息量相当大，数字化后要占用大量的存储空间，传统的存储设备（如磁带、磁盘等）根本无法满足这一要求。光存储技术的发展为存储多媒体信息提供了保证，是解决计算机存储容量问题的重要突破。

光盘存储技术是 20 世纪 80 年代存储技术领域最重大的发明之一，该项技术在 20 世纪 90 年代得到了广泛的应用，对多媒体信息存储技术的发展产生了深远的影响。近年来，光盘生产已形成具有相当规模的新兴产业，光盘存储技术及其应用已进入成熟期，其物理尺寸、编码方式、数据记录方式及数据文件的组织方式都有了国际标准，并朝着高密度、大容量、小体积、多品种、快速存取及网络化的方向发展。

9.1 光盘存储系统

光盘存储系统由光盘和光盘驱动器组成。作为一种新兴的信息存储手段，光盘存储技术在计算机外部存储设备应用上得到了飞速的发展。它已向磁盘存储技术提出挑战，在许多新的应用领域展示了强大的生命力。目前，磁盘存储在存取时间和数据传输率上占有优势，光盘如果在这两方面赶上磁盘存储技术，将会有更大的作为。在光盘存储技术上，材料和器件是光存储发展的主流，而提高光存储密度和数据传输率是发展光存储的主攻方向。

9.1.1 光盘及其特点



图 9.1 光盘外形

CD (Compact Disc) 意为高密盘，称为光盘，如图 9.1 所示，它是通过光学方式来记录和读取二进制信息的。光盘是一种数字式记录存储器，具有一切存储介质的优点，如大容量、耐用、易保存、标准化等。

光盘在存储多媒体信息方面主要有以下特点。

1. 记录密度高

由于使用相干性好的激光作为能源，可把它聚焦成直径约 $1\mu\text{m}$ 的光点进行记录，存储一位信息所需的介质面积仅约为 $1\mu\text{m}^2$ ，因而存储密度可高达 $10^7\sim 10^8\text{b}/\text{cm}^2$ ，为目前磁盘的数十倍至上百倍。

2. 存储容量大

一张标准的 CD-ROM 光盘容量可达 650MB，相当于 450 张 3.5 英寸软盘的容量，如果用来存放文字，则可以存储相当于 16 万张 A4 纸的信息量，足以容纳 800 多部 40 万字的书籍，如图 9.2 所示。例如著名的 Compton 交互式百科全书，在一张 CD-ROM 上就包含了来自印刷版的 3500 篇文章（900 万词），7000 多张照片和插图，以及 14 个小时的声音和 80 多个

录像片段。从 1995 年开始，国外光盘百科全书的销售量已经超过了用纸印刷的百科全书销量，人类获取知识和经验的途径从此发生了根本性的变化。

正因为 CD-ROM 容量巨大，除了大量用于电子出版物外，也将其作为软件发行的载体，如 Microsoft 的 Windows 系统都以光盘形式提供给用户。

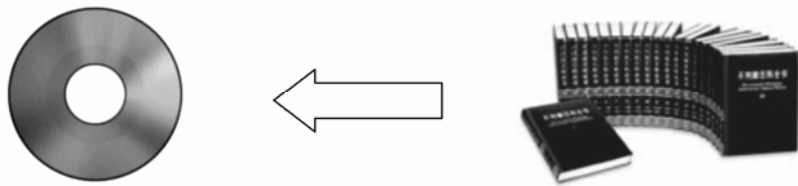


图 9.2 一张光盘可存放 800 多部 40 万字的书籍

3. 采用非接触方式读/写信息

这是光盘存储技术所具有的独特性能。在读取光盘信息时，光盘与光学读/写头不互相接触。这样的读/写方式当然就不会使盘面磨损、划伤，也不会损伤激光头。此外，光盘的记录层上附有透明的保护层，记录层上不会产生伤痕和灰尘。光盘外表面上的灰尘颗粒与划伤，对记录信息的影响很小。

4. 信息保存时间长

对于只读型光盘，不必担心文件会被误删除，更不必忧虑受到病毒的侵扰。如果使用得当，一张光盘上的信息可保存长达几十年甚至更长。

5. 不同平台可以互换

CD-ROM 盘片上的信息按照 ISO 9660 标准格式记录，即使在不同的硬件或软件平台上，CD-ROM 中的信息也可以被正确读出。

6. 多种媒体融合

光盘可以同时存储文字、图形、图像、声音等信息媒体。以光盘为介质的各种电子出版物目前已十分普及，它内容丰富、图文并茂、引人入胜，大大增加了读者的阅读兴趣，而且还易于将信息按相关性进行组织，以方便用户引用。

7. 价格低廉

与磁带和磁盘相比，光盘是目前计算机数据最便宜的存储介质。

9.1.2 光盘的类型

按照数据存放格式和类型，光盘可分为许多不同的类型，并以不同的名称以示区别。因而光盘通常是指下列光盘的总称：CD-DA，CD-ROM，CD-R，CD-RW，VCD，DVD 等。

按光盘的读/写性能来讲，可分为 3 种类型。

1. 只读型光盘存储器

只读光盘中的数据是用压模方法压制而成的，用户只能读取上面的数据，而不能写入或修改光盘中的数据。它适用于大量的、通常不需要改变的数据信息存储，如各类电子出版物、大型软件的载体。

最常见的只读光盘为 CD-ROM 光盘，当然，CD-DA，VCD，DVD 等也都属于只读式光盘。

2. 可记录光盘存储器

这种光盘是一种写入后不能修改但允许反复多次读出（Write Once, Read Multiple）的光盘存储器，主要用于重要数据的长期保存。目前得到了广泛应用的 CD-R 就属于这类光盘。

CD-R 是英文 CD Recordable 的简写。CD-R 信息的写入系统主要由写入器和写入控制软件构成。写入器也称为光盘刻录机，是写入系统的硬件部分，写入控制软件也称为光盘刻录软件，如 Nero Burning Rom，Copy To CD/DVD 等。

CD-R 数据读出的原理与 CD-ROM 相同，因而 CD-R 盘片上刻录的数据既可以在 CD-R 刻录机上也可以在普通的 CD-ROM 驱动器中读取。

CD-R 光盘适用于存储数据、文字、图形、图像、声音等多种媒体，并且具有存储可靠性高、寿命长和检索方便等突出优点，得到极为广泛的应用。特别是为那些需要永久性存储信息而不准擦除或更改的用户提供了一种最佳方案。

3. 可擦写光盘存储器

可记录式光盘的不足之处是写入数据后不允许修改，操作过程一旦有误则可能导致整张盘片报废。可擦写光盘具有磁盘一样的可擦写性，可多次写入或修改光盘上的数据，更适合作为计算机的新型标准外存设备。

CD-RW（CD-Rewritable）是可擦写光盘最具有发展前途的代表，允许用户存储、移动、修改光盘上的数据。可擦写光盘的推出具有划时代的意义，它将改变人们使用光盘的方式。

9.1.3 光盘驱动器

必须使用专门的设备才能从光盘中读取数据，这种设备就是光盘驱动器，简称光驱，如图 9.3 所示。

CD-ROM 光盘驱动器既能读 CD-ROM 盘，也能读其他如 CD-DA，CD-R 和 VCD 等光盘，但它不能读 DVD 等特殊形式的光盘。这些光盘各有自己的标准，需要相应的光驱才能访问。因此在谈到光盘驱动器时，要注意说明它对哪些光盘是适用的。

性能指标参数是生产厂商在产品推出过程中的标称值，包括接口类型、数据传输率、平均寻道时间、内部数据缓冲、多种光碟格式支持等。

1. 接口类型

光驱的接口方式有两种：IDE 接口和 SCSI 接口。IDE 是目前普遍使用的光驱接口方式，具有安装方便、价格便宜的特点。SCSI 接口的光驱价格相对要贵得多，需要 SCSI 接口卡支持，安装也较烦琐，但具有稳定、CPU 耗用低的优点，因而，网络服务器广泛采用的都是 SCSI 接口的光驱。

2. 速度和读取方式

光驱的读取速度以 150KB/s 数据传输速率的单倍速为基准，如 24 速光驱其数据传输速率即为 3600KB/s。但由于数据读取方式的限制，高倍速光驱并不能总是运行在其标称的速度下，只是在读取某一位置时才达到最大的数据传输率。

光盘信息的组织和地址编码与软盘、硬盘不一样，它的光道是螺旋形的，如图 9.4 所示，光驱在工作时有 CLV 和 CAV 两种方式。



图 9.3 光盘驱动器

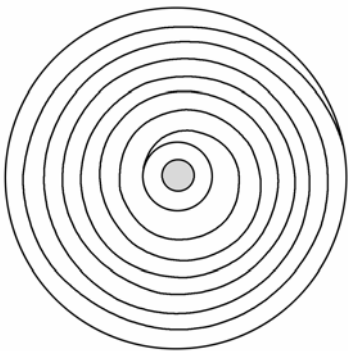


图 9.4 光盘螺旋形光道

在恒定线速度（Constant Line Velocity, CLV）方式下，单位距离的光道上所存储的信息容量是相等的，即内外光道的数据记录密度相同，因而可以充分利用盘片的空间，增加了数据存储容量。但这样一来，激光头每旋转一圈所读取的数据量就不一样了，内圈数据少，外圈数据多。因此在 CLV 方式下，当激光头移动到不同的轨道时，电机也必须以不同的速度旋转，内圈转得快些，外圈转得慢些，以维持单位距离信息读取时间一致。对于高速运转的光驱来讲，CLV 方式容易造成光驱耐用性的降低。

目前的高倍速光驱大多采用了恒定角速度（Constant Angular Velocity, CAV）技术。在 CAV 方式下，不管是内圈还是外圈，激光头始终以恒定的角速度旋转光盘盘片，这和硬盘驱动器的操作方式很相似。恒定的转速对电机来说比较容易实现，由于不需要在随机寻道时经常地改变电机转速，因此随机读取性能得到很大的改善。不过，在 CAV 方式下，光盘内、外圈转动的线速度是不相等的，因此内、外光道的数据记录密度也是不同的，光盘的存储空间没有充分利用。正是由于 CLV 与 CAV 技术各有优劣，于是一些光驱采用 CLV+CAV 技术，在内圈采用 CLV 方式，以保证足够的传输速率；在外圈采用 CAV 方式，以提高可靠性。

3. 容错能力

任何光驱的性能指标中都没有标出容错能力的参数，但这却是一个实在的光驱评判标准。在高倍速光驱设计中，高速旋转的电机使激光头在读取数据的准确定位性能上相对于低倍速光驱要逊色许多，同时劣质的光碟更加剧对光驱容错能力的需求，因而许多厂家都加强对容错能力的设计。一些小厂家只是单纯加大激光头的发射功率，初期使用时读盘容错能力非常好，但在两三个月之后，激光头老化，其容错性能明显下降。而名牌大厂通常以提高光驱的整体性能为出发点，采用先进的机芯电路设计，可改善数据读取过程中的准确性和稳定性，或者根据光碟数据类型自动调整读取速度，以达到容错纠错的目的。因此在选择光驱时

除了要有较好的容错能力外，还要注意其整体性能的优良。必须注意的是，为了保证数据读取的严密性，光驱产品不可能具有同 VCD 影碟机一样的超强纠错能力，两者设计的出发点和使用目的都不相同。


4. 缓存和格式支持

光驱本身所带的缓存在一定程度上能够提高数据传输效率，缓存越大在发送数据给 PC 之前就能准备或存储更大的数据段。一般来说，驱动器速度越快，就需要更大容量的缓存，以处理更高的传输速率。现在光驱已将缓存提高到 2MB，读盘速度、盘片接入时间都得到了显著的提升。缓存的提高还能有效防止停顿和马赛克，保证读盘的准确、稳定和流畅性。另外随着可擦写光盘驱动器的普及，对包括 CD-R/RW，DVD 盘片在内的多种光碟类型的支持也显得非常重要，这无疑扩大了光驱作为多媒体部件的使用范围。

5. 其他

光驱高速旋转的主轴电机带来的震动、噪声、发热对光盘有一定的影响，选择有防震机构、静噪性能的产品对光驱和光盘都有好处。

【例 9.1】 明基 652A（增强型银色）的性能参数如图 9.5 所示。



光驱种类：CD-ROM

安装方式：内置

CD-ROM 读取倍速：52X

接口类型：ATAPI-IDE

缓存大小：128KB

放置特点：水平/垂直

图 9.5 明基 652A（增强型银色）的性能参数

9.2 光盘的标准

由于光盘能存储不同类型的数据，包括音频和视频数据、计算机程序等，而这些数据的组织方式各不相同，因此制定了一些国际标准，以适应多媒体的各种应用。标准对各类光盘的物理尺寸、编码方式、数据记录方式及数据文件的组织方式都有详细的规定。目前主要的光盘标准及产品有 CD-DA，CD-ROM，CD-R，VCD，DVD 等。

光盘技术发展过程中的主要产品如图 9.6 所示。

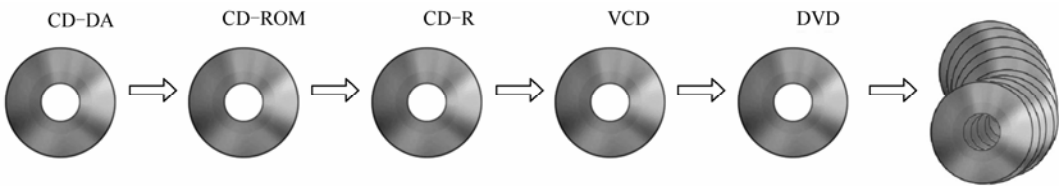


图 9.6 光盘技术发展过程中的主要产品

9.2.1 CD-DA标准（红皮书标准）

CD-DA(Compact Disc-Digital Audio),称之为数字式激光唱盘或CD唱盘。1982年,SONY和PHILIPS公司合作开发了数字光盘音响系统,并制定了CD唱盘标准。因为该标准出版时采用了红色封面,因此也称为红皮书标准。CD唱盘是CD家族的第一个成员,其标准是以后各种新的CD标准的基础。CD-DA用于存储数字化的高保真立体声音乐。

9.2.2 CD-ROM标准（黄皮书标准）

该标准从CD-DA发展而来,又称黄皮书标准,该标准使得光盘以统一的格式存储只读的信息。CD-ROM是在CD-DA之后产生的,尽管两者之间有许多相似之处,但是它们有一个根本的区别:音频CD只能存放音乐,而CD-ROM可以存放文本、图形、声音、视频及动画。CD-ROM中存放的是计算机数据,对误码率有一定的要求,所以在CD-DA的基础上又增加了一层错误检测和错误校正。

黄皮书只定义了计算机数据在光盘上的物理存储的格式,但计算机数据在外存储器上都是以文件形式存储的,而黄皮书并未定义CD-ROM的逻辑结构,即文件结构。为这个问题,国际标准化组织(ISO)于1988年制定了ISO 9660标准,这是一个为CD-ROM盘上的文件系统而定义的标准,它定义了CD-ROM光盘上通用的信息交换的文件规范,该规范由于其通用性和相对的独立性,ISO 9660很快成为CD-ROM光盘上最为重要的文件格式。它的出现,对CD-ROM的普及应用产生了很大的推动作用。从理论上讲,只要按照此标准规范,借助于相应的驱动程序,CD-ROM上的文件在不同的平台都可以读出。ISO 9660文件系统适用的平台包括DOS, Windows, UNIX, Linux, Macintosh, OS/2等。

9.2.3 CD-R标准（橙皮书标准）

CD-R(Compact Disk Recordable,可记录光盘)是一种可以单次写入,多次读取的光盘,如图9.7所示。CD-R标准是PHILIPS和SONY公司于1990年制定的,被称做橙皮书Part。基于橙皮书的CD-R空白光盘实际上没有记录任何信息,一旦按照某种文件格式并通过刻写程序和设备,将需要长期保存的数据写入空白的CD-R盘片上,这时的CD-R空白盘就可以变成CD-DA, CD-ROM或VCD光盘的形式。

CD-R盘与CD-ROM盘相比有许多共同之处,它们的主要差别在于CD-R盘上增加了一层有机染料作为记录层。当写入激光束聚焦到记录层上时,染料被加热后烧熔,形成一系列代表信息的凹坑。这些凹坑与CD-ROM盘上的凹坑类似,但CD-ROM盘上的凹坑是用金属压模压出的。

由于光盘成本上的差别,CD-R一般不适于大量制作的产品,但在少量制作时它呈现出许多优越性。例如,CD-R的制作成本低,制作一张光盘的费用基本上就是材料费,省去了传统工艺中母盘制作的大量开销,是数据永久备份的廉价手段;CD-R也可以多次写入数据,不必一次将盘写满;CD-R还可以作为提供给生产厂家用来大批量生产CD盘片的样品母盘。CD-R的另外一个重要特性是它可以像其他CD光盘一样在普通的CD-ROM驱动器上使用。

CD-RW(Compact Disk Rewritable,可擦写光盘)代表一种“重复写入”的技术,利用这种技术可以在特殊光盘上的相同位置重复写入数据。1996年初,Ricoh、PHILIPS、SONY、Yamaha、Hewlett-Packard及Mitsubishi化学公司建立了一个工业论坛,发布了CD-RW格式

标准。该标准由 Ricoh 大力发展，并于 1996 年推出了业界第一款 CD-RW 驱动器，型号为 MP6200S，是一种刻录速度 2 倍速，2 倍速复写，6 倍速读取的产品。同时橙皮书 Part III 也正式出台，从此 CD-RW（CD-Rewritable）标准被正式定义。



图 9.7 CD-R、CD-RW 光盘

从那时开始 CD-RW 驱动器就取代了 CD-R 驱动器，CD-RW 驱动器完全兼容 CD-R 驱动器，完全支持 CD-R 盘片。CD-RW 光盘刻录的方式与 CD-R 光盘相同，区别就在于其可以擦除并多次重写。这样可以将 CD-RW 盘视做软盘，可以进行文件的复制、删除等操作，方便灵活。

CD-RW 光盘与 CD-R 光盘主要有 4 个方面不同：可重写、价格更高、写入速度慢和反射率更低，CD-RW 盘片只有性能优越的 CD-ROM 光驱才能读出。

【例 9.2】 明基 5232X CD 刻录机的性能参数如图 9.8 所示。



光驱种类：CD 刻录机
安装方式：内置
刻录机规格：CD-RW
CD-RW 擦写倍速：32X
CD-R 写入倍速：52X
CD-ROM 读取倍速：52X
写入方式：Disc-at-once, Track-at-once, Session-at-once, Multisession, Packet writing, Raw writing
缓存大小：2MB

图 9.8 明基 5232X CD 刻录机的性能参数

9.2.4 Video-CD标准（白皮书标准）

Video-CD（简称 VCD）是由 JVC，PHILIPS 等公司于 1993 年联合制定的数字电视存储技术规范，称之为白皮书（White Book）。它用来描述光盘上存放采用 MPEG—1（活动图像专家组）标准编码的全动态图像及其相应立体声声数据的光盘格式。

VCD 的出现与 MPEG—1 标准有着密切的关系。MPEG—1 是一个专用于处理活动影像的标准，也是一个与特定应用对象无关的通用标准。VCD 按照 MPEG—1 标准对音、视频数据进行压缩后，提高了存储空间的有效利用率，使一张盘片能存放 74 分钟的活动图像与伴音。SVCD 是 VCD 的改进标准，它采用 MPEG—2 压缩，采用可变压缩率来获得较好的视频质量。VCD 被广泛使用，VCD 和 DVD 在发行市场上已完全取代录像带和镭射影碟的地位。VCD 被认为会被 DVD 所取代，但是由于它的一些特点，还将在市场上保持一段时间。VCD 没有像 DVD 一样的区码限制，这意味着它可以在任何兼容机器上观看。有些节目因为成本问题，不会制作 DVD 或 VHS 录像带版本，购买者只能购买 VCD 版本。VCD 比 DVD 廉价，制作成本也较低。

VCD 标准在 ISO 9660 文件结构基础上定义了完整的文件系统，这样就使 VCD 节目能够在 CD-ROM 和 VCD 播放机上播放。VCD 盘需要的目录有 Root Directory 0（根目录 0）、CDI、VCD 和 MPEGAV。VCD 盘的文件目录结构如图 9.9 所示。Video CD 对文件目录做了如下规定。

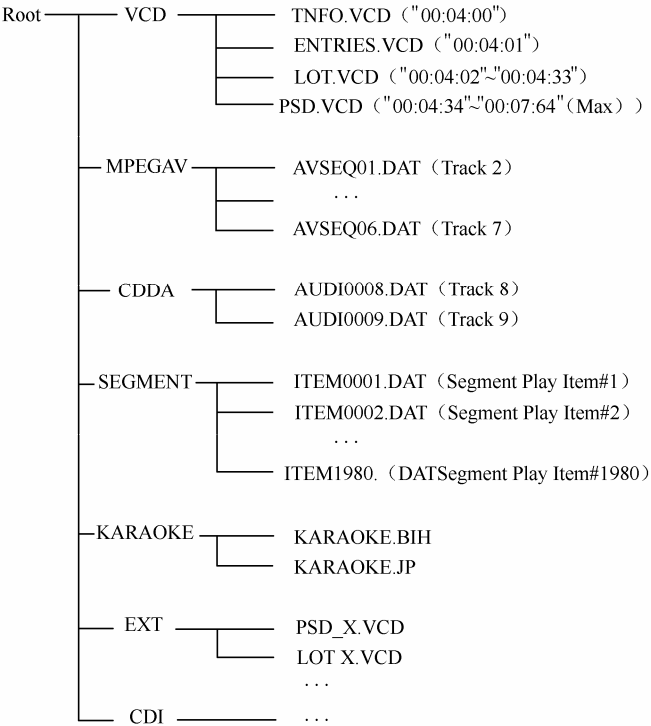


图 9.9 VCD 的文件目录结构

- ① 如果 VCD 盘包含卡拉 OK 基本信息区（Karaoke Basic Information Area），这个区域中的文件必须存放在 KARAOKE 目录下。
- ② 如果 VCD 盘包含分段播放项目区（Segment Play Item Area），这个区域中的文件必须存放在 SEGMENT 目录下。
- ③ 如果有扩展的播放顺序描述符（Play Sequence Descriptor，PSD）文件，这个文件必须存放在 EXT 目录下。
- ④ VCD 信息区（Video CD Information Area）中的文件必须存放在 VCD 目录下。
- ⑤ 所有表示 MPEG Audio/Video 光道的文件都必须存放在 MPEGAV 目录下。
- ⑥ 所有表示 CD-DA 光道的文件都必须存放在 CDDA 目录下。

9.2.5 DVD标准

DVD 在其发展初期，是指 Digital Video Disc，即“数字视盘”。随着光盘技术的进步，它不仅可以存储影视节目，而且可以存储音乐、数据，随着用途增多，就将这种类型的光盘统称为“数字多用途光盘”，英文名是 Digital Versatile Disc，简称 DVD。

大家知道，MPEG—1 的图像是家用录像机级的质量，MPEG—1 技术的成熟促成了 VCD 的诞生、产业的形成和市场的成熟。而 MPEG—2 的视频是广播级的图像质量，由于广播级数字电视的数据量要比 MPEG—1 的数据量大得多，而 CD 的容量尽管有 650MB 至 700MB，但仍满足不了存储 MPEG—2 视频节目的要求。为了解决 MPEG—2 视频节目的存储问题，就促成了 DVD 的问世。

在 1995 年，一个由 SONY 和 PHILIPS Electronics DV 公司领导的国际财团与另一个由 Toshiba 和 Time Warner Entertainment 公司领导的国际财团分别提出了两个不兼容的高密度 CD（High Density Compact Disc，HDCD）规格。同年 10 月，两大财团终于同意盘片的设计按 Toshiba/Time Warner 公司的方案，而存储在盘上的数据编码则按 SONY/PHILIPS 公司的方

案。最终的单面单层 DVD 盘片应该能够存储 4.7GB 的数据，存储 133min 的 MPEG—2 Video，并配备 Dolby AC3/MPEG—2 Audio 质量的声音，支持多达 8 种不同语言的伴音和 32 个字幕通道。



图 9.10 DVD 标志

DVD 的标志如图 9.10 所示。
下面分别说明 DVD 的分类。

1. 按格式分类

DVD 有 6 种主要格式，它的标准一般用 Book X（X= A，B，C，...）来表示。

- ① Book A: DVD-ROM，只读光盘，用途类似 CD-ROM。
- ② Book B: DVD-Video，影视光盘，用途类似 LD 或 Video CD。
- ③ Book C: DVD-Audio，音乐光盘，用途类似音乐 CD。
- ④ Book D: DVD-R，一次写入的光盘，用途类似 CD-R。
- ⑤ Book E: DVD-RAM，可擦写的光盘。
- ⑥ Book F: DVD-RW，可擦写的光盘，用途类似 CD-RW。

DVD 分类的标志如图 9.11 所示。



图 9.11 DVD 分类标志图

需要说明的是，现在还有两种采用 DVD 盘片结构的可录光盘：DVD+R 和 DVD+RW，这是 DVD 联盟制定的格式。

2. 按盘面大小分类

DVD 目前有 120mm 和 80mm 两种规格。

3. 按刻录方式分类

DVD 光盘有只读光盘和刻录光盘两种。

(1) 只读光盘

只读光盘有 DVD-ROM, DVD-Video, DVD-Audio3 种格式。

① DVD-ROM。

DVD-ROM 是数据型光盘, 它的容量是 CD-ROM 容量的 7 倍, 数据传输率则是 CD-ROM 的 9 倍。它被广泛应用于计算机领域。DVD-ROM 的文件系统采用 UDF 和 ISO 9660 标准, 与计算机的操作系统兼容, 它在数据存取、多媒体、计算机游戏方面有广泛的应用。

② DVD-Video。

DVD-Video 俗称 DVD 影碟, 是 DVD 的视频格式。

③ DVD-Audio。

DVD-Audio 的规格是 DVD 音频格式。它可选择 16, 20, 24 位量化; 可选择 48/96/192kHz 和 44.1/88.2/176.4kHz 采样频率; 可选择 2 声道或多声道, 最多可有 6 个声道。

如果选择最高量化 24 位, 最高采样频率 192kHz, 录制 2 声道立体声节目, 那么, 一张 DVD-5 光盘可以存储 64 分钟这样高品质的音频节目。CD-DA 是 16 位量化, 44.1kHz 采样, 它的数据量仅为 DVD-Audio 的 15%。

DVD-Audio 采用的 LPCM 是未经压缩的原音重现。它也可以采用杜比数字、DTS 规格。

(2) 刻录光盘

刻录光盘分为一次写入和可擦写两种类型, 如图 9.12 所示。目前 DVD 刻录技术有 3 大类、5 种规范 (DVD-RAM, DVD-R, DVD-RW, DVD+R, DVD+RW)。一次写入型有 DVD-R, DVD+R; 可擦写型有 DVD-RAM, DVD-RW, DVD+RW。

DVD-R, DVD-RW 和 DVD-RAM 是 DVD 论坛 (DVD Forum, DVD 论坛是 DVD 格式持有者共同创办的一个组织, 目的是建立“单一、最佳”的 DVD 格式, 促进 DVD 产品的普及) 制定的格式。



图 9.12 DVD 刻录盘

DVD+R 和 DVD+RW 是 DVD 联盟 (从 DVD 论坛中分裂出来的若干家厂商, 并有戴尔、惠普等加入) 制定的格式。

目前市场上许多光盘刻录机能够兼容上述大部分或全部规范（如 SuperMulti 类产品），价格也比较合理，有些还能支持双层刻录。

4. 按记录层分类

DVD 光盘可分为 5 种类型：单面单层 DVD-5、单面双层 DVD-9、双面单层 DVD-10、双面双层 DVD-18 和双面混合层 DVD-14。DVD-14 是将单层记录面和双层记录面黏合而成的光盘。

表 9.1 列出了 DVD 各种光盘的情况，可以与 CD-ROM 比较。

表 9.1 按记录层分类 DVD 类型

类 别	外 径	光 盘 种 类	容量（GB）
只读 DVD	120mm	DVD-5 单面单层	4.70
		DVD-9 单面双层	8.54
		DVD-10 双面单层	9.40
		DVD-14 双面混合层	13.24
		DVD-18 双面双层	17.08
	80mm	DVD1 单面单层	1.46
		DVD2 单面双层	2.66
		DVD3 双面单层	2.92
		DVD4 双面双层	5.32
刻录 DVD	120mm	DVD-R V1.0 单面单层	3.95
		DVD-R V2.0 单面单层/双面单层	4.70/9.40
		DVD-RW V2.0 单面单层/双面单层	4.70/9.40
		DVD-RAM V1.0 单面单层	2.58
		DVD-RAM V1.0 双面单层	5.16
		DVD-RAM V2.0 单面单层	4.70
		DVD-RAM V2.0 双面单层	9.40
	80mm	DVD-R V1.0 单面单层	1.2
		DVD-RAM V1.0 单面单层	0.7
		DVD-RAM V2.0 双面单层	1.46
CD-ROM	120mm	CD-ROM	0.650
	80mm	CD-ROM	0.194

DVD 具有高存储密度、高画质、高音质、高兼容性和高可靠性等特点。

1) 高存储密度

从外观和尺寸方面来看，DVD 盘与现在广泛使用的 CD 盘没有什么差别，直径均为 120mm，厚度 1.2mm（DVD 是由两片 0.6mm 的盘片黏合而成的）。但 DVD 的存储容量比 CD 大得多，普通的单面单层 DVD 容量是 4.7GB，相当于 CD 的 7 倍，而双面双层的 DVD 容量可达到 17GB，相当于 CD 的 25 倍。DVD 与 CD 技术参数比较见表 9.2。

表 9.2 DVD 与 CD 技术参数比较

技 术 参 数	DVD	CD
盘片直径	120mm	120mm
盘片厚度	0.6mm+0.6mm	1.2mm
激光波长	635/650nm	780nm
光轨间距	0.74μm	1.6μm
最小光坑长度	0.4μm	0.83μm
纠错码	RSPC	CIRC
通道码	8-16 调制	8-14 调制，加 3 位合并位
盘片表面的利用率	87.5cm ²	86.0cm ²

DVD 容量比 CD 大的主要原因如下。

(1) 更高的存储密度

DVD 盘光道之间的间距由 CD 光盘的 1.6μm 缩小到 0.74μm，而记录信息的最小凹、凸坑长度由原来的 0.83μm 缩小到 0.4μm，这是 DVD 盘的存储容量可提高到 4.7GB 的主要原因，它的容量是 CD 的 7 倍，它们之间的差别如图 9.13 所示。为了把光道距离和信息记录凹、凸坑的长度和宽度做得更小，DVD 利用波长更短、聚焦更细的红色激光进行信息的存取，这是因为光学读出头的分辨率和激光波长成正比。

(2) 更有效的编码

DVD 信号的调制方式和错误校正方法也做了相应的修正以适应高密度的需要，CD 存储器采用 8-14（EFM）加 3 位合并位的调制方式，而 DVD 则采用效率比较高的 8-16+（EFM PLUS）的方式，这是为了能够和 CD 光盘兼容；CD 光盘采用的错误校正系统是里德-索洛蒙码（Cross-Interleaved Read-Solomon Code，CIRC），而 DVD 采用里德-索洛蒙乘积码（Reed Solomon Product Code，RSPC），它比 CIRC 更可靠。

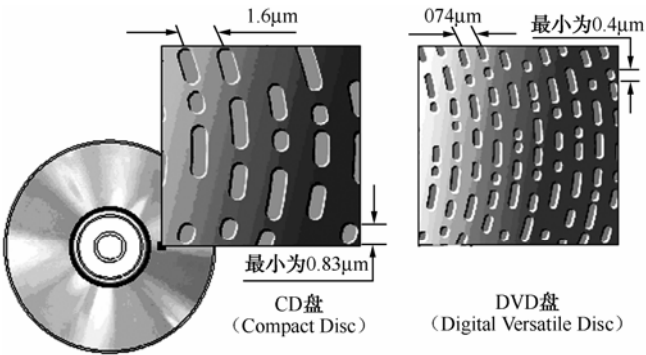


图 9.13 DVD 比 CD 有更高的存储密度

采用修改的数据编码和调制算法都可以减少 DVD 盘上的冗余位，从而为用户提供更多的存储空间。现在的 CD 需要用 17 位来表示一个 8 位的数据（14 个通道位和 3 个用于改善读出信号的合并位）。新的算法将使用 16 位来表示一个 8 位的数据，这样也增加了 DVD 的容量。

此外，在 CD 盘上有许多 EDC（Error Detection Code）和 ECC（Error Correction Code）信息位，采用新的算法之后这些信息位的数目可以减小，也就相当于增加用户数据的容量。

（3）增加盘的数据记录面积

加大盘的数据记录区域也是提高记录容量的有效途径。DVD 盘的记录区域从 CD 盘的 86cm^2 提高到 86.6cm^2 ，如图 9.14 所示。

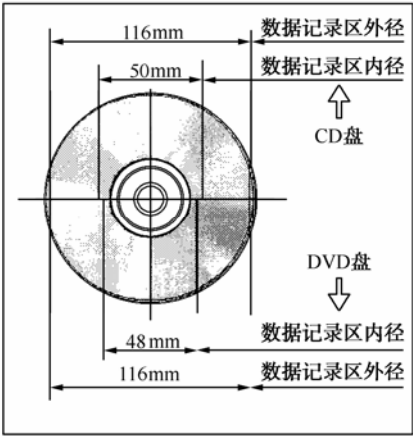


图 9.14 增加盘的数据记录面积

（4）增加记录面、记录层

提高 DVD 存储容量的另一个重要措施是使用盘片的两个面来记录数据，以及在一个面上制作好几个记录层，这无疑会大大增加 DVD 盘的容量。单面双层 DVD 盘的容量可达到 8.5GB，而双面双层 DVD 盘的容量可达到 17GB。

2) 高画质

DVD 采用国际通用的活动图像压缩标准 MPEG—2（ISO/IEC 13818），其系统码率是可变的(1~10.7Mbps)。目前 DVD-Video 选用 MPEG—2 的 11 种规范中的主型主级规范 MP@ML（即 NTSC 制式电视 720 像素/行×480 行/帧，30fps；PAL 制式电视 720 像素/行×576 行/帧，25fps，数据传输速率最大为 15Mbps），其图像质量达到广播级电视水平，其水平分辨率为 500 线。要实现更高清晰度的画质，还可选用 MPEG—2 中对应的高级规范。DVD 的系统码流由主视频码流（MPEG—2/MPEG—1 压缩码流）、子图像码流（最多可录放 32 个码流，用于 32 种文字电影对白和卡拉 OK 字幕显示）和声频码流（最多可录放 8 个码流，支持 8 种语言声音）3 部分组成。整个系统码流的最大数据速率可达 10.08Mbps。DVD 还具有多结局（欣赏不同的多种故事情节发展）、多角度（多达 9 个角度观看图像）和父母控制（又称儿童锁，屏蔽儿童不宜观看的画面）等功能。画面的长宽比可选择全景扫描、4：3 普通屏幕和 16：9 宽屏幕等方式。

3) 高音质

DVD 音频系统必选方式是线性 PCM（LPCM）和杜比数字（Dolby Digital，DD）等，此外，还可以选 DTS 等格式。表 9.3 列出了 DVD 视频光盘音频编码技术参数。

表 9.3 DVD 视频光盘音频编码技术参数

必 选 方 式	NTSC 视频光盘必选以下两种方式 ①LPCM（立体声）（未压缩） ②杜比数字（1～5.1 声道）（压缩方式）	PAL/SECAM 视频光盘必选以下两种方式 ①LPCM（立体声）（未压缩） ②下列压缩方式之一： 杜比数字（1～5.1 声道），MPEG—2 Audio
任 选 方 式	5.1 声道 杜比数字 环绕声（Dolby Digital Surround Sound） 5.1 声道 DTS 环绕声 5.1 声道 MPEG—2 环绕声	

LPCM(线性脉冲编码调制)技术是将未经压缩的双声道信号利用 PCM 格式存储在 DVD 盘片上，由于没有经过任何有损音质的数据压缩处理，因此可以获得最佳的音质表现。这种 PCM 格式原理跟 CD 唱片是一样的，但由于 DVD 具有更大的空间与数据读取速度优势，可以使用更高的采样频率和量化精度。现在，要求高音质的 DVD 采用 96kHz/24 位，使 DVD 影碟的音质优于 CD 唱片。

杜比数字 DD (Dolby Digital)，也称杜比 AC3，是杜比公司的一项专利编码技术。杜比数字是一项先进的感觉编码技术，用于传输和存储多达五个全频带声道（前左、前右、中置、左后、右后）的信息，另加一个低频效果声道。杜比数字技术被广泛应用于电影和 DVD 光盘中，它还是制定于 1998 年的美国新清晰度电视系统（ATSC）的音频标准，并于 1999 年成为欧洲数字电视标准（DVB）的可选音频制式。在 DVD 光盘中，杜比数字可以是单声道，双声道，也可以是 5.1 声道，这要根据制片商的要求来设置。杜比数字标志如图 9.15 所示。



图 9.15 杜比数字的标志

DTS (Digital Theater Systems) 来源于电影院音响系统，是美国 DTS Technology 公司开发的一种数字立体环绕音响制式。DTS 采用的压缩方式有别于杜比数字，典型的压缩比为 3：1，仅为杜比数字的四分之一。其次，DTS 的 96kHz 的采样频率为杜比数字 48kHz 的一倍，并轻而易举地将数据传输率提高到 1536kbps，比杜比数字的 448kbps 高出了 3 倍，这使得 DTS 在声音的细节表现、空间感方面更加优秀，不过目前它还远不如杜比 AC3 流行。DTS 标志如图 9.16 所示。



图 9.16 DTS 的标志

4) 高兼容性

DVD 向下兼容。DVD 光驱可以读取 CD 数据，DVD 播放机能播放 VCD 碟片和 CD 唱片，DVD 刻录机也能刻录 CD 光盘。

目前主要有两类 DVD 驱动器，一类是专门用于播放 DVD 影碟的 DVD 影碟机，另一类是安装在 PC 上使用的 DVD-ROM。就驱动器而言，两者的结构和原理是一样的，与 CD 光盘驱动器也无本质的区别。与普通 CD 光盘 1 倍速 150KB/s 的传输速率不同，DVD 的 1 倍速数据传输速率达到 1350KB/s，大约是普通 CD-ROM 的 9 倍。对于播放 DVD 影碟 1 倍速就够了，虽然更高的数据传输率可以让播放影片时扫描更平滑，搜索数据更快，但不能从本质上改变 DVD 影片的品质。而读取数据类型的 DVD（如游戏光盘），传输速率是越高越好。驱动器中最关键的部件是激光头，由于 DVD 与 CD 使用的激光波长不同，为了使 DVD 向下兼

容 CD/VCD, DVD 激光头的设计比 CD 更复杂。还有一种将 CD-RW 刻录机和 DVD-ROM 驱动器组合在一起的一体式光驱（称为 CD-RW/DVD-ROM COMBO, 康宝），由于具有多种功能，该产品已经逐步推广使用。

【例 9.3】 三星金将军 52X 康宝 2M（白金版）性能参数如图 9.17 所示。



图 9.17 三星金将军 52X 康宝 2M（白金版）性能参数

5) 高可靠性

DVD 采用 RSPC (Reed-Solomon Product Code, 里德-所罗门乘积码) 纠错编码方式和 8~16 位信号调制方式, 确保数据读取可靠。若原始误码率为 10^{-3} , 经纠错后, 误码率可小于 10^{-20} , 远远低于计算机所需的误码率 10^{-12} 。为了有效地防止软件被复制, 在美国活动图像协会 (Motion Picture Association of America) 的积极参与下, 于 1996 年 7 月同东芝、索尼等 12 家家电与计算机公司就 DVD 软件版权与防盗版问题达成协议。1996 年 10 月, 由各方组成的 DVD 技术联合会公布了 DVD 软件和硬件采用的乱码技术, 以及按 6 个地区区域码分区发行软件的措施, 实现了软件著作权保护与可靠使用。

6 个区域码划分如下。

第 1 区 (Region 1): 美国、加拿大、东太平洋岛屿。

第 2 区 (Region 2): 日本、西欧、北欧、南非、中东 (包括埃及)。

第 3 区 (Region 3): 中国台湾和中国香港地区, 韩国、泰国、印尼等东南亚国家。

第 4 区 (Region 4): 澳洲, 新西兰, 中南美洲, 南太平洋岛屿, 加勒比海。

第 5 区 (Region 5): 俄罗斯、蒙古、印度半岛、中亚、东欧、朝鲜、北非、西北亚一带等。

第 6 区 (Region 6): 中国大陆。

6 个区域码的标志如图 9.18 所示。



图 9.18 DVD 6 个区域码标志

DVD 论坛要求 DVD 的碟片和播放机都要设置区域码，播放机只能播放相同区域码的碟片。坚持要使用区域码识别，主要是因为每部电影在世界各地上映时间不同，为避免电影未上演，DVD-Video 却先行上市造成电影业界的利益损失，就以区域码加以控制，使其他地区的影碟不能在本地地区播放。

区域码保护了版权，但对使用者来说也带来了不便。目前只有美国八大电影公司出版的 DVD 影碟才有严格的限制，其他的 DVD 碟片一般没有区域限制，称为自由区域码（Free Region Code）或称全区片。其标志如图 9.19 所示。

DVD 区域码的限制可以分为两种方式，一种称为 RPC2（RPC，Region Protection Control），是由硬件来控制区域码，DVD-ROM 硬件制造厂商在 DVD-ROM 生产时就在光驱的 BIOS 中限定用户只能更改 5 次区域码，所以在使用时必须注意选择适合的 DVD 影片，当区域码修改超过 5 次后，将无法再更改区域码，除非送回原制造厂商刷新。而另一种称为 RPC1，这时硬件没有区域码的限制，而是由软件来控制。如 PowerDVD 允许使用者可做 5 次区域码修改，此后，PowerDVD 就会锁在你最后一次播放的影片区域码，而不能再做区域码的更改，除非重新安装操作系统。



图 9.19 DVD 影片全区码标志

习 题 9

1. 填空题

- (1) 光盘信息的组织和地址编码与软盘、硬盘不一样，它的光道是螺旋形的，光驱在工作时有_____和_____两种方式。
- (2) 光驱的读取速度以单倍速为基准，24 速 CD-ROM 光驱的数据传输率为_____，24 速 DVD-ROM 光驱的数据传输率为_____。
- (3) 视频图像标准，VCD 采用的是_____标准，DVD 采用的是_____标准。
- (4) 光盘中最基本的数据存储单元是_____。
- (5) _____标准规定了光盘文件结构、目录结构和路径表。
- (6) _____区段是一般的刻录方式，这种方式比较省空间，压制的光盘大多数是这种刻录方式。

2. 简答题

- (1) 相对于其他存储介质，光盘在存储多媒体信息方面具有哪些主要特点？
- (2) 常见的光盘类型有哪几种？各适用于哪些场合？
- (3) CLV 和 CAV 技术有什么不同？

- (4) 光盘驱动器的性能指标有哪些?
- (5) DVD 光盘类型有哪些?
- (6) DVD 的主要特点有哪些?
- (7) 为什么 DVD 比 CD 的容量大很多?
- (8) DVD 在音频和视频处理上, 采用了哪些技术?
- (9) CD-ROM, CD-R, CD-RW, DVD 的结构及存取信息的原理是什么?
- (10) 盘面上标明“74 分钟, 650MB”的 CD-R 光盘, 刻满后其容量一定一样吗?

第 10 章 网络多媒体技术

多媒体技术与网络技术的有机结合满足了信息化社会人们对各种信息的大量需求。网络多媒体技术的迅速发展，加速了多种网络多媒体技术的应用，如视频会议系统、数字视频监控系統、多媒体电子邮件、视频点播（VOD）、远程多媒体数据库等。

10.1 超文本与超媒体

在当今信息社会，信息以爆炸的方式不断增加，而且种类不断增长，除了文本、数字之外，图形、图像、视频、音频等多媒体信息已在信息处理领域中占有越来越大的比重。如何对多媒体化的海量信息进行存储和检索，以及像人类思维那样通过“联想”来明确不同信息块之间的连接关系和相似性，已成为重要课题。超文本/超媒体技术的出现，为我们实现多媒体信息综合有效的管理带来了希望，尤其在 Internet 飞速发展的今天，超文本/超媒体技术成为 Internet 信息检索的核心技术。

10.1.1 超文本/超媒体的基本概念

超文本是一种文本，它和书本上的文本是一样的。但传统的文本文件相比，它们之间的主要差别是，传统文本是以线性方式组织的，而超文本是以非线性方式组织的。这里的“非线性”是指文本中遇到的一些相关内容通过链接组织在一起，用户可以很方便地浏览这些相关内容。这种文本的组织方式与人们的思维方式和工作方式比较接近，如图 10.1 所示。

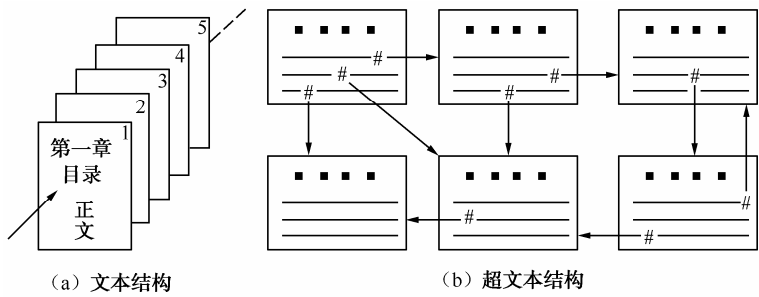


图 10.1 文本结构与超文本结构

超文本可以简单地定义为收集、存储和浏览离散信息，以及建立和表示信息之间关系的技术。超文本可引用、链接其他不同类型（内含声音、图片、动画等）的文件，这些具有多媒体操作的超文本，称为超媒体（Hypermedia），即超媒体=超文本+多媒体。在 WWW 上引入超媒体技术，实现了语音和图像的实时获取，为我们展现了丰富多彩的世界。

10.1.2 超文本/超媒体组成要素

超文本/超媒体是由相对独立的节点信息和表达它们之间关系的链所组成的信息网络，所

以，节点、链和网络是超文本的三要素。

1. 节点

节点是超文本/超媒体表达信息的基本单位，它是描述某个特殊主题的数据集合。节点中可以定义链与其他节点相连接。实际上，超文本/超媒体网络是由具有复杂的链接结构的节点所组成的。节点内容可以是文本、图形、图像、视频、音频等，也可以是它们的组合形式。

2. 链

链用来链接相关的节点，也是组成超文本/超媒体的基本单位，是超文本/超媒体的灵魂。链形式上是从一个节点指向另一个节点的指针，本质上表示不同节点间存在着信息联系。在超文本/超媒体系统中，定义的各节点可用链连接起来，链的功能强弱，直接影响节点信息的表现力，也影响网络的结构。

链的一般结构分为链源、链宿及链的属性。链源是链的起始端，是导致节点信息迁移的原因；链源是文本节点中的字词，称为热字；链源是图像上的某一敏感区，称为热区；视频节点的一个特性是时间连续性，节点的链源与时间有关，即视频的某几帧，称为热点。链宿是链的目的地，一般是节点。链的属性决定链的类型。

3. 网络

超文本/超媒体中由节点和链构成网络信息获取的路径，通过交互式的浏览、查询来获取信息。由此可见，超文本/超媒体的信息网络在建立信息联系的同时，也表现出作者的思维轨迹。它不仅提供了知识、信息，而且还蕴涵着对它们的分析、推理等，其内涵极为丰富。

用户在浏览较大型超文本/超媒体系统时，很容易迷失方向，不知道自己在超文本/超媒体网络中处于什么位置，也不知道下一步该沿哪条链继续浏览。所以，超文本/超媒体系统必须要有强有力的导航工具，才能查看浏览线路，判断当前节点方位。目前解决迷路问题的方法主要包括用户界面方法、结构化方法、搜索/查询方法等。如用户界面方法，通过改善用户界面来帮助导航。它将超文本/超媒体数据库中的节点及节点之间关系的部分或全部以图形方式显示出来，使用户对节点内容、节点之间关系和系统的整体轮廓有清晰的认识，如图 10.2 所示。

10.1.3 超文本/超媒体的实现机制——HTML语言

超文本/超媒体的思想需要一种网络的实现机制，这就是 HTML 语言。1989 年 3 月，在欧洲粒子物理实验室（CERN），富于想象力的物理学家 Tim Berners-Lee 提出了一项计划，目的是使科学家们能很容易地翻阅同行们的文章。为了支持此计划，Tim 创建了一种新的语言来传输和呈现超文本文档。这种语言就是超文本标记语言（Hyper Text Markup Language，HTML）。

HTML 提供了将声、文、图结合在一起，综合表达信息的强有力的手段。HTML 利用统一资源定位器（URL）来表示超链接，并在文本内指向其他网络资源。它能使用户将文档中的词和图像与其他文档链接起来，而不论这些文档存放在何处，用户只需在具有链接的文字或图像上单击一下，就可以将 Internet 中与其相关联的有关信息查找出来并显示在屏幕上。

用 HTML 可描述文本、图形、图像及声音等多媒体元素的显示格式及链接，用户可方便地在关联的链接之间自由选择，快速获取自己感兴趣的内容，如图 10.3 所示。

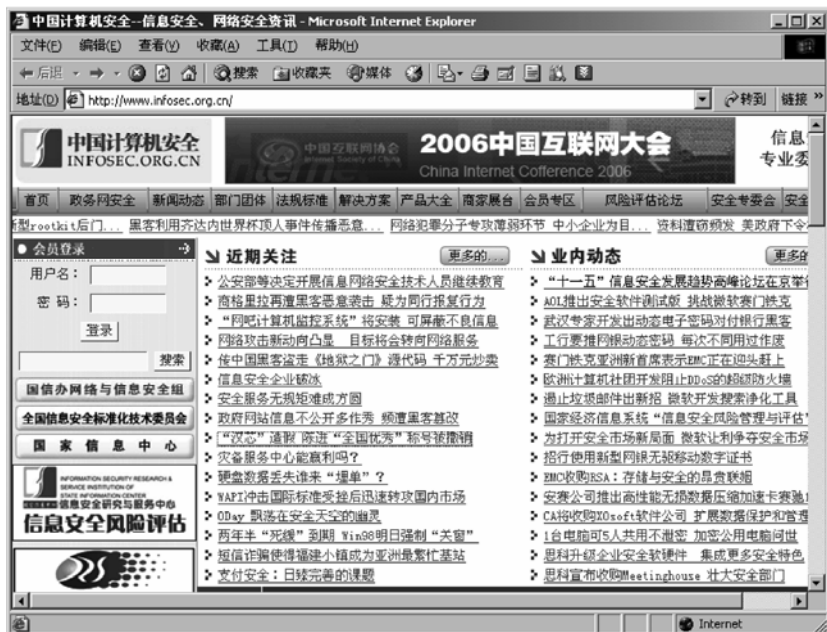


图 10.2 网页上的超级链接

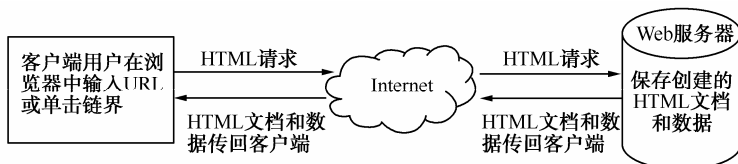


图 10.3 HTML 文档在链接时的传送过程

HTML 对不同的媒体信息使用标记 (Tag) 来控制以达到预期的显示效果。HTML 的标记按用途可以分为以下不同的类型。

- 基本标记：用于创建一个 HTML 文档，设置文档的标题及文档的可见部分。
- 标题标记：设置标题栏中的文档标题。
- 文档整体属性标记：设置文档的背景、文字颜色、各类超级链接的颜色。
- 文本标记：设置文本的字体、字号和文字颜色等。
- 链接标记：创建内部或外部的超级链接。
- 格式排版标记：设置文档段落的格式。
- 图像元素标记：在文档中添加图像并设置图像的位置、边框、显示的图像大小等。
- 表格标记：创建表格并设置表头的格式。
- 表格属性标记：设置表格的大小、对齐方式、边框等表格属性。
- 框架标记：定义框架的大小，以及在不支持框架的浏览器中显示的提示。
- 框架属性标记：设置框架的内容、边框、滚动条，以及是否允许用户调整窗口。
- 表单标记：用于创建表单、滚动菜单、下拉菜单、文本框、单/复选框、按钮等。

一般来讲，HTML 的元素有下列 3 种表示方法。

- ① <标记名>文本</标记名>
- ② <标记名 属性名=“属性值”>文本</标记名>
- ③ <标记名>

第①种语法结构适用于基本标记、标题标记、文本标记和表格标记等。

第②种语法结构适用于文档属性标记、字体设置标记、超级链接标记、图形元素标记、表格属性标记、窗框标记和表单属性等。

第③种语法仅适用于一些特殊的元素，比如分段元素 p，其作用是通知 WWW 浏览器在此处分段，因而不需要界定作用范围，所以它没有结尾标注。为保持语法上的严谨，在 HTML 3.0 标准中，也定义了</p> 标注，它用于需要界定作用范围的段落，比如增加对齐方式属性的段落。

【例 10.1】 下面是一段 HTML 代码：

```
<html>
<head>
<title>
This is a example!
</title>
</head>
<body bgcolor=#6699FF>
<h1 align="center"><font face="隶书">静夜思</font> </h1>
<h3 align="center"><font face="隶书">床前明月光，疑是地上霜。</font></h3><p>
<h3 align="center"><font face="隶书">举头望明月，低头思故乡。</font></h3><p>
</body>
</html>
```

由上面的代码可见一个基本的 HTML 文件有两部分：文件“头”（head）指定了一个标题，当网页被显示的时候它就会在浏览器的标题栏上出现；文件“体”（body）包含了文本、图像和链接等的信息。需要注意的是：对于 HTML 文件中的标记来说，英文字母的大小写不做区分，如<title>和<TITLE>或者<TiTIE>是一样的。显示效果如图 10.4 所示。

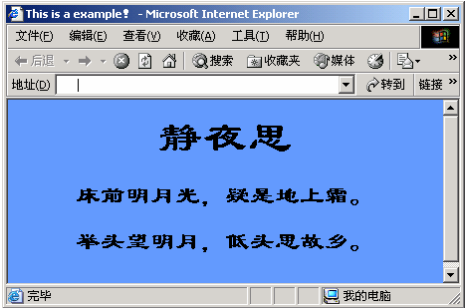


图 10.4 HTML 网页效果

【例 10.2】 在网页中插入图像，代码如下：

```
<html>
<head>
<title>
This is a example!
</title>
</head>
<body >
<p align="center"></p>
</body>
</html>
```

其中 src="lena.jpg"中的 lena.jpg 是由用户自己指定的一幅图像，具体位置（可以是网络地址或本机地址）、文件名与内容由用户自行设定。保存代码并运行后显示效果如图 10.5 所示。

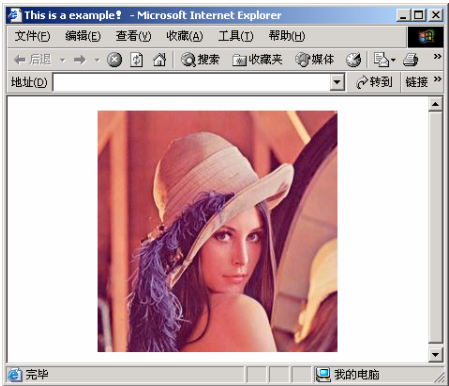


图 10.5 在网页中插入图像

【例 10.3】 在网页中插入声音，代码如下：

```
<html>
<head>
<title>
This is a example!
</title>
</head>
<body >
<p align="center">
<embedsrc="file:///c:/WINNT/Media/ringout.wav"width=350height=40type=audio/x-pn-realaudio
-plugin AutoStart="true" Loop= "true"></embed></p>
</body>
</html>
```

其中，“file:///c:/WINNT/Media/ringout.wav”是由用户指定的，具体位置（可以是网络地址或本机地址）、文件名与内容由用户自行设定。保存代码并运行便可听到播放的音乐，播放控制各按钮均可由用户调节，如图 10.6 所示。

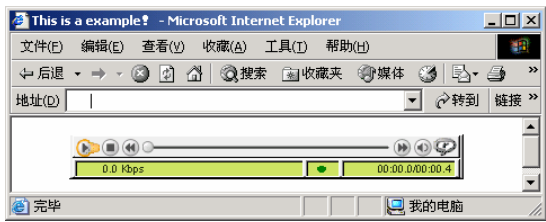


图 10.6 在网页中插入音乐

【例 10.4】 在网页中插入视频，代码如下：

```
<html>
<head>
<title>
This is a example!
</title>
</head>
<body >
<p align="center">
<embedsrc="file:///c:/WINNT/clock.avi" width=300 height=300 type=audio/x-pn-realaudio-plugin
AutoStart="true" Loop= "true"></embed></p>
</body>
</html>
```

其中，“file:///c:/WINNT/clock.avi”是由用户指定的，具体位置（可以是网络地址或本机地址）、文件名与内容由用户自行设定。保存代码并运行便可看到视频的内容，播放控制各按钮均可由用户调节，如图 10.7 所示。

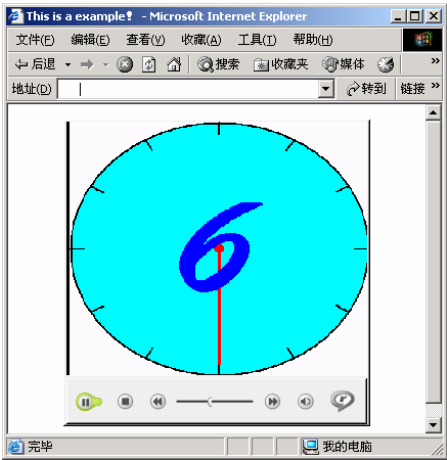


图 10.7 在网页中插入视频

一些常用的 HTML 标记及其功能见表 10.1。

表 10.1 HTML 中的常用标记及其功能

标 记 类 型	标 记 名	标记的功能描述
基本标记	<html></html>	创建一个 HTML 文档
	<head></head>	设置文档标题及其他不在 Web 网页上显示的信息
	<body></body>	设置文档的可见部分
标题标记	<title></title>	设置文档标题栏中显示的标题
文档整体 属性标记	<body bgcolor=?>	设置背景颜色，使用名字或十六进制值
	<body text=?>	设置文本文字颜色，使用名字或十六进制值
	<body link=?>	设置链接颜色，使用名字或十六进制值
	<body vlink=?>	设置已使用的链接的颜色，使用名字或十六进制值
	<body alink=?>	设置正被击中的链接颜色，使用名字或十六进制值
文本标记	<pre></pre>	创建预格式化文本
	<h1></h1>	创建最大的标题
	<h6></h6>	创建最小的标题
		创建黑体字
	<i></i>	创建斜体字
	<tt></tt>	创建打字机风格的字体
	<cite></cite>	创建一个引用，通常是斜体
		加重一个单词（通常是斜体加黑体）
		设置字体大小，从 1~7
		设置字体的颜色，使用名字或十六进制值
链接标记		创建一个超级链接
		创建一个自动发送电子邮件的链接
		创建一个位于文档内部的靶位
		创建一个指向位于文档内部靶位的链接
排版格式 标记	<p>	创建一个新的段落
	<p align=?>	将段落按左、中、右对齐
	 	插入一个回车换行符
	<blockquote></blockquote>	从两边缩进文本
	<dl></dl>	创建一个定义列表
		创建一个标有数字的列表
		在每个数字列表项之前加上一个数字
		创建一个标有圆点的列表
		在每个圆点列表项之前加上一个圆点
	<div align=?>	用于对大块 HTML 段落排版，也用于格式化表

标 记 类 型	标 记 名	标记的功能描述
图像元素 标记		添加一个图像
		排列对齐一个图像：左、中、右或上、中、下
		设置围绕一个图像的边框的大小
	<hr>	加入一条水平线
	<hr size=?>	设置水平线的大小（高度）
	<hr width=?>	设置水平线的宽度（百分比或绝对像素点）
	<hr noshade>	创建一个没有阴影的水平线
表格标记	<table></table>	创建一个表格
	<th></th>	设置表格头：一个通常使用黑体居中文字的单元格
表格属性 标记	<table border=#>	设置围绕表格的边框的宽度
	<table cellspacing=#>	设置表格单元格之间空间的大小
	<table cellpadding=#>	设置表格单元格边框与其内部内容之间空间的大小
	<table width=# or %>	设置表格的宽度：像素值或文档总宽度的百分比
	<tr align=?>or<td align=?>	设置表格单元格的水平对齐方式（左、中、右）
	<tr valign=?>or<td valign=?>	设置表格单元格的垂直对齐方式（上、中、下）
	<td colspan=#>	设置一个表格单元格应跨占的列数（默认值为 1）
	<td rowspan=#>	设置一个表格单元格应跨占的行数（默认值为 1）
	<td nowrap>	禁止表格单元格内的内容自动换行
框架标记	<frameset></frameset>	创建一个框架，它可以嵌在其他框架文档中
	<frameset rows="value,value">	定义框架的行数，可以用绝对像素值或高度的百分比
	<frameset cols="value,value">	定义框架的列数，可以用绝对像素值或宽度的百分比
	<frame>	定义一个框架内的单一窗口或窗口区域
	<noframes></noframes>	定义在不支持框架的浏览器中显示的提示信息
框架属性 标记	<frame src="URL">	设置框架内显示的 HTML 文档
	<frame name="name">	命名框架或区域以便别的框架可以指向它
	<frame marginwidth=#>	定义框架左右边缘的空白大小，必须大于等于 1
	<frame marginheight=#>	定义框架上下边缘的空白大小，必须大于等于 1
	<frame scrolling=VALUE>	设置框架是否有滚动栏，可取 yes, no 及 auto（默认值）
	<frame noresize>	禁止用户调整一个框架的大小
表单标记	<form></form>	创建表单
	<select multiple name = "NAME" size =?></select>	创建一个滚动菜单，size 设置在滚动前用户可以看到的事项数目
	<option>	设置每个表单项或菜单项的内容
	<select name="NAME"> </select>	创建一个下拉菜单

标 记 类 型	标 记 名	标记的功能描述
表单标记	<textarea name="NAME" cols=? rows=?></textarea>	创建一个文本框，cols 的值为文本框的宽度，rows 的值为文本框的高度
	<inputtype="checkbox"name="NAME">	创建一个复选框，文字在标记后面
	<inputtype="radio"name= "NAME" value="x">	创建一个单选框，文字在标记后面
	<inputtype=txtname="NAME" size=20>	创建一个单行文本输入区域，size 设置以字符计的宽度
	<inputtype="submit"value="NAME">	创建一个 submit（提交）按钮，NAME 为按钮名称
	<input type="image" border=0 name="NAME"src="name.gif">	创建一个使用图像的 submit（提交）按钮
	<input type="reset">	创建一个 reset（重置）按钮

用 HTML 组织的文件本身属于普通的文本文件，它可以用一般的文字编辑软件编辑，如记事本、Microsoft Word 等。也可以使用专门的 HTML 文件编辑软件来编辑，如 Microsoft 公司的 FrontPage，Macromedia 公司的 Dreamweaver 等。HTML 文件的扩展名可以是.html 或.htm，现有的网页浏览器都支持这两种类型的 HTML 文件。

10.2 多媒体通信网络

随着多媒体技术的发展及多媒体应用的不断深化，大量数字化的音频和视频信息需要统一的信息网络来传输。通过高速网络可实现大量的数字化数据处理、交换和通信，以达到相互间的共享。

现有的许多通信网络的设计目的多样，多数已得到广泛的应用，包括电话交换网、Ethernet、FDDI、分组交换网、ISDN、VOD、HFC 等，它们分别属于电信网、计算机网和有线电视网。这些网络之间已存在不同程度的交叉与融合，但是要使这些不同的网络统一起来还为时过早。

10.2.1 多媒体通信的发展背景

多媒体计算机技术的崛起是多媒体通信发展的首要原因。20 世纪 80 年代初，美国、日本和欧洲著名的计算机公司开始致力于多媒体技术的研究，并把该技术应用于 PC。首先建立了基于局域网（LAN）的多媒体通信系统。自 20 世纪 90 年代开始，多媒体计算机技术就成为计算机领域的热点之一。计算机在各个领域中的广泛应用使得人类可以获取的信息爆炸性地增长，当技术发展到可以方便地处理各种感觉媒体时，多媒体计算机技术便自然而然地出现并迅速发展起来。多媒体通信中的“多媒体”一词，指的是由在内容上相互关联的文本、图形、图像、音频和视频等媒体数据构成的一种复合信息实体。计算机以数字化的方式对任何一种媒体进行表示、存储、传输和处理，并且将这些不同类型的媒体数据有机地合成在一起，

形成多媒体数据，这就是多媒体计算机技术。多媒体计算机技术综合和发展了计算机科学中的多种技术，如操作系统、计算机通信、数字信号和图像处理等。它是以计算机为核心的，集图、文、声、像处理技术为一体的综合性处理技术。随着科学技术的迅速发展和社会需求的日益增长，人们已不满足于单一媒体提供的传统的单一服务，如电话、电视、传真等，而是需要诸如数据、文本、图形、图像、音频和视频等多种媒体信息以超越时空限制的集中方式作为一个整体呈现在人们的眼前。在这种时代背景下，伴随着多媒体计算机技术与电话、广播、电视、微波、卫星通信、广域网和局域网等各种通信技术相结合，产生了一种边缘性技术——多媒体通信。

10.2.2 多媒体通信的特点

多媒体通信（Multimedia Communications）是多媒体技术与通信技术的完美结合，它突破了计算机、通信、电子等传统领域的界限，把计算机的交互性、通信网络的分布性和多媒体信息的综合性融为一体，多媒体对通信的影响主要表现在以下几个方面。

（1）多媒体通信数据量巨大

由于多媒体数据量很大，存储空间要求大，传输带宽要求高，因此不可避免地要对所传输的数据进行压缩。而现在的高倍率的压缩以损失原始数据信息量为代价，这影响到媒体本身的质量。在很多情况下，不得不采用静态、慢速或小画面等办法来限制数据量，这也影响通信质量。因此，真正实现多媒体通信，必须加大带宽，使得通信网络能适应多媒体数据量的增长。

（2）多媒体通信的实时性

多媒体中的声音、动画、视频等媒体对多媒体传输设备的要求很高，即使带宽充足，如果通信协议不合适，也会影响多媒体数据的实时性。例如，在语音通信时，偶尔的误码不去纠正的效果要比由于纠错重发而发生语音停顿的效果要好得多。一般来说，电路交换方式延时短，但占用专门信道，不易共享；而分组交换方式则延时偏长，且不适于数据量变化大的业务使用。很显然，这将要求通信网、通信协议及高层协议能适应这种需求。

实时性的影响还存在于端到端的延迟上，在多媒体数据传输中，许多处理环节都会增加端到端的延迟。鉴于各种多媒体之间的特性如此不一致，一般采用“服务质量（Quality of Service, QoS）”来描述，传输时也往往根据 QoS 来决定传输策略。例如，对语音可采取延迟短、延迟变化小的传输策略，对数据传输则可采用可靠、保序的传输策略等。

（3）多媒体通信的同步性

同步性指的是在多媒体通信终端上显现的图像、声音和文字是以同步方式工作的。例如，用户要检索一个重要的历史事件的片段，该事件的运动图像（或静止图像）存放在图像数据库中，其文字叙述和语言说明放在其他数据库中。多媒体通信终端通过不同传输途径将所需要的信息从不同的数据库中提取出来，并将这些声音、图像、文字同步起来，构成一个整体的信息呈现在用户面前，使声音、图像、文字实现同步，并将同步的信息送给用户。

同步性是多媒体通信系统最主要的特点之一。信息的同步与否，决定了系统是多媒体通信系统还是多种媒体通信系统。此外，多媒体通信的同步性也是最为困难的技术问题之一。一般来说，多媒体通信系统是一个资源受限的系统，所谓的资源受限有两种情况，即通信速率受限和终端内存受限。如果这两个方面没有限制，同步本来不会有很大的技术难点。例如，如果信道通信速率不受限，那么只要发送端完全安排好信息媒体间的关系，在接收端又完全

忠实地再现出来，信息同步将不成问题，当然在信道的通信速率受限的情况下，接收端的信息同步就要困难得多。另外，如果接收端存储器的存储量是无限的，将所有信息全部接收下来，然后在终端内同步播出，这种场合下同步问题也好解决，但实际上这个条件经常是无法满足的，因而使同步问题变得很困难。

（4）多媒体通信的交互性

多媒体系统的关键特点是交互性。这就要求多媒体通信网络提供双向的数据传输能力，这种双向传输通道从功能和带宽来讲都是不对称的。

（5）分布式处理和协同工作

目前的通信网络状况是多网共存，在未来的通信系统中，多网统一、业务综合和多媒体化应是发展的重点。现有的各类信息网络，包括电话网、计算机网，甚至电视网，广播网和新型信息网将集成为一个网络，不同的业务在其上运行，以一个插口、一个号码和一个体系面对用户。为了达到这个目标，在高速宽带的网络上，实现各种多媒体信息的传输就非常必要了。

分布式处理是向用户提供综合服务的基本方法。因为多媒体引入到了分布式处理领域后，不仅是通信传输的问题，还有许多建立在通信传输之上的分布式处理与应用问题需要研究。需要解决诸如各项多媒体应用在分布式环境下运行时，如何通过分布式环境解决多点、多人合作问题，以及如何提供远程的多媒体信息服务等问题。

10.2.3 多媒体通信的关键技术

多媒体通信的关键技术主要有：

- ① 声音、视频、图像等多媒体信息处理技术；
- ② 数据压缩和解压缩技术；
- ③ 多媒体信息实时传输与同步技术；
- ④ 多媒体通信协议与标准化。

10.3 网络多媒体应用

高速网络技术的发展，大大改善了网络的多媒体应用环境，推动了网络多媒体应用的发展，出现了很多网络多媒体应用系统，如可视电话、多媒体会议系统、多媒体邮件系统、多媒体信息咨询系统、交互式信息点播系统、远程教育系统、远程医疗系统，IP 电话等。同时网络多媒体应用的发展也对计算机网络技术、数据存储技术和分布式处理技术等提出了更高的要求，带动了相关技术的进步。下面将介绍几种典型的网络多媒体应用系统，即视频会议系统、视频点播系统、IP 电话，从中可以看出这些系统的不同技术特色和风格。

10.3.1 网络多媒体应用概述

多年来，国际电信联盟（ITU）为公共和私营电信组织制定了许多多媒体计算和通信系统的推荐标准，以促进各国之间的电信合作。ITU 的 26 个（Series A~Z）系列推荐标准中，与多媒体通信关系最密切的 7 个系列标准见表 10.2，3 种类型网络多媒体应用系统的核心技术标准见表 10.3。

表 10.2 ITU 系列与多媒体通信关系最密切的 7 个标准

系 列 名	主 要 内 容
Series G	传输系统、媒体数字系统和网络
Series H	视听和多媒体系统
Series I	综合业务数字网（ISDN）
Series J	电视、声音节目和其他多媒体信号的传输
Series Q	电话交换和控制信号传输
Series T	远程信息处理业务的终端设备
Series V	电话网上的数据通信

表 10.3 3 种类型网络多媒体应用系统的核心技术标准

项 目	H.320	H.323（V1/V2）	H.324
发布时间	1990	1996/1998	1996
应用范围	窄带 ISDN	带宽无保证分组交换网络	PSTN
图像编码	H.261, H.263	H.261, H.263	H.261, H.263
声音编码	G.711, G.722, G.728	G.711, G.722, G.728,G.723.1, G.729	G.723.1
多路复合控制	H.221, H.230/H.242	H.225.0, H.245	H.223, H.245
多点协议	H.231, H.243	H.323	
实时通信	T.120	T.120	T.120

20 世纪 90 年代初开发的电视会议标准是 H.320，它定义了通信的建立、数字电视图像和声音压缩编码的算法，运行在综合业务数字网（Integrated Services Digital Network，ISDN）上。在 56kbps 传输率的通信信道上支持帧速率比较低的电视图像，而在 1.544Mbps 传输率的信道（即 T1 信道）上可以传输 CIF 格式的满帧速率电视图像。在局域网上的桌面电视会议（Desktop Video Conferencing）采用 H.323 标准，这是基于分组交换的网络多媒体应用系统。在公众交换电话网（Public Switched Telephone Network，PSTN）上的网络桌面电视会议使用调制解调器，采用 H.324 标准。因特网上的电视会议目前大部分都趋向于采用 H.323 标准和正在开发的 SIP 标准，使用 IP 协议提供局域网上的电视会议，而全球的因特网电视会议目前还不能保证实时电视会议的服务质量。

在多媒体通信标准中，电视图像的编码标准都采用 H.261 和 H.263。H.261 主要用来支持电视会议和可视电话，并于 1992 年开始应用于综合业务数字网络（ISDN）。该标准采用帧内压缩和帧间压缩技术，可使用硬件或者软件来执行。电视图像数据压缩后的数据速率为 $P \times 64\text{kbps}$ ，其中， P 的变动范围为 1~30，这取决于所使用的 ISDN 通道数。H.261 支持 CIF 和 QCIF 的分辨率。H.263 是在 H.261 的基础上开发的电视图像编码标准，用于低位速率通信的电视图像编码，目标是改善在调制解调器上传输的图像质量，并增加了对电视图像格式的支持。

计算机网络是网络多媒体应用的基础，电路交换网络与分组交换网络的融合是构造网络多媒体应用系统结构的出发点。如图 10.8 所示为网络多媒体应用系统的结构示意图。从图中

可以看到，网络多媒体应用系统主要由下面几个部件组成：网关（Gateway）、会务器（Gatekeeper）和通信终端（通信终端包括执行 H.320，H.323 或 H.324 协议的计算机和执行 H.324 的电话机）。此外，H.323 还定义了一个叫做多点控制单元（Multipoint Control Unit，MCU）的部件，它是 H.320 和 H.323 的一个重要设备，可作为一个单独的设备接入到网络上，但现在开发的一些产品则把它要实现的功能集成到会务器中，因此图中未画出。在 H.323 协议中，把通信终端、网关、会务器或者 MCU 叫做端点（Endpoint）。

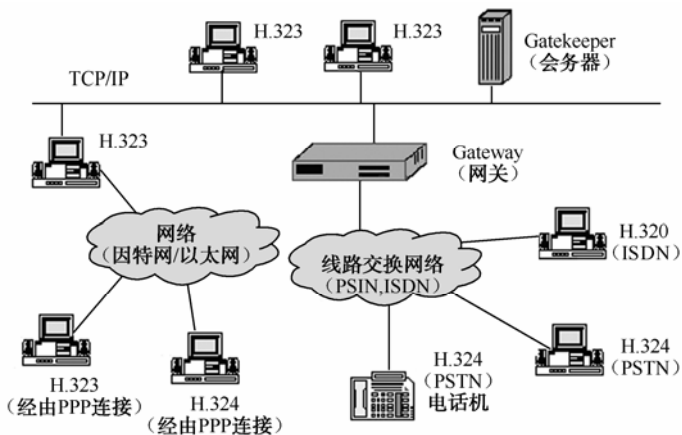


图 10.8 网络多媒体应用系统的结构示意图

网关和会务器是网络多媒体应用系统的两个极其重要的组成部件。网关提供面向媒体的功能，如传送声音及电视图像数据和接收数据包等。会务器提供面向服务的功能，如身份验证、呼叫路由选择和地址转换等。网关和会务器密切配合完成多媒体通信的任务。

10.3.2 视频会议系统

视频会议系统的结构如图 10.9 所示，它主要由视频会议终端、多点控制器、信道（网络）及控制管理软件组成。

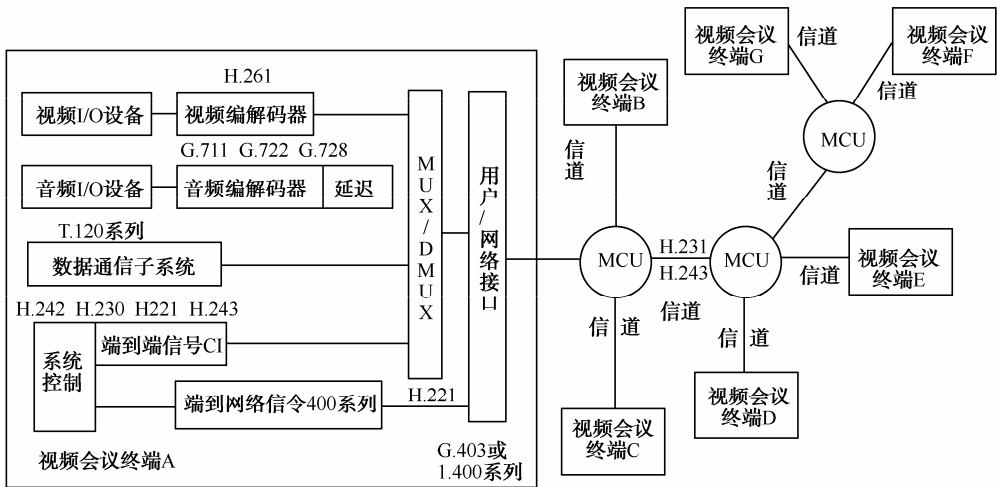


图 10.9 视频会议系统结构框图

视频会议系统终端的主要功能是：完成视频信号的采集、编辑处理及显示输出、音频信号的采集、编辑处理及输出、视频音频数字信号的压缩编码和解码，最后将符合国际标准的压缩码流经线路接口送到信道，或从信道上将标准压缩码流经线路接口送到终端中。此外，终端还要形成通信的各种控制信息：同步控制和指示信号、远端摄像机的控制协议、定义帧结构、呼叫规程及多个终端的呼叫规程、加密标准、传送密钥及密钥的管理标准等。

多点控制单元 MCU (Multipoint Control Unit) 是视频会议系统的关键设备，它的主要功能是对视频、语音及数据信号进行切换，例如它会把传送到 MCU 某会场发言者的图像信号切换到所有会场。对于语音信号，若同时有几个发言，可以对它们进行混合处理，选出最高的音频信号，切换到其他会场。MCU 的主要组成部分是：网络接口单元、呼叫控制单元、多路复用和解复用单元、音频处理器、视频处理器、数据处理器、控制处理器、密钥处理分发表及呼叫控制处理器。

视频会议系统的服务质量 QoS (Quality of Service) 是满足视频会议系统需求的核心问题，视频会议系统要把用户的服务请求映射成预先规定的 QoS 参数，进而与系统和网络资源对应起来，通过资源的分配和调度满足用户的应用需要。资源的分配和调度可以选用资源的静态管理和动态管理去完成。资源的静态管理包括：QoS 的协商和解释、资源许可 (Admission)、资源的保留和分配及资源的释放。资源的动态管理包括：进程管理、缓冲区管理、传输率和流量控制及差错控制。

视频会议系统最后一个组成部分是安全保密系统，它也是视频会议的一个重要问题。安全保密系统的主要组成部分是加密模块和解密模块，加密模块是将会议终端的用户数据加密形成加密后的数据在网络上传输，解密模块接收加密数据进行解密得到用户数据。加密和解密模块的核心是密钥的生成和管理，密钥生成的核心是加密算法，加密算法不包含在国际标准的建议中，它由视频会议系统设计者研制或选用。

根据通信节点的数量，视频会议系统可以分为：点对点视频会议系统和多点视频会议系统。

点对点视频会议系统支持两个通信节点间视频会议通信功能，它的主要业务是：

① 可视电话。可视电话是现有公用电话网上使用的具有双工视频传送功能的电话设备。

由于电话网带宽的限制，可视电话只能使用较小的屏幕和较低的视频帧率。例如使用 3.3 英寸的液晶屏幕，每秒钟可传送 2~10 帧画面。

② 桌面视频会议系统。这种视频会议系统利用用户现有的台式机 (如 PC) 平台以及网络通信设备和远地另一台装备了同样或兼容设备的台式机通过网络进行通信，这种系统仅限于两个用户或两个小组用户使用。Intel 公司的 Proshare Personal Conferencing Video System200 是这类系统的一个典型示例，这是一种点对点的个人视频会议系统，支持 ISDN 和 LAN 的连接，采用硬件编码压缩，软件解压缩，为了方便协同工作，Proshare 还提供共享笔记本和共享应用程序。

③ 会议室型视频会议系统。在会议室型视频会议系统的支持下，一群与会者集中在一间特殊装备的会议室中，这种会议室作为视频会议的一个收发中心，能与远地的另外一套类似的会议室进行交互通信，完成两点间的视频会议功能。由于会议室与会者较多，因此对视听效果要求较高，一套典型的系统一般应包括：一台或两台大屏幕监视器、高质量摄像机、高分辨率的专用图形摄像机，复杂的音响设备，控制设备及其他可选设备，以满足不同用户的要求。

多点视频会议系统：多点视频会议系统允许三个或三个以上不同地点的参加者同时参与会议。多点视频会议系统的一个关键技术是多点控制问题，多点控制单元（MCU）在通信网络上控制各个点的视频、音频、通用数据和控制信号的流向，使与会者可以接收到相应的视频、音频等信息，维持会议正常进行。

10.3.3 视频点播系统

1. VOD系统模型

视频点播（Video On Demand，VOD）系统是一种交互式多媒体信息服务系统，用户可根据自己的需要和兴趣选择多媒体信息内容，并控制其播放过程。这种新的多媒体信息服务形式被广泛应用于有线电视系统、远程教育系统，以及各种公共信息咨询和服务系统中。VOD系统采用 C/S（Client/Server）模型，如图 10.10 所示。它主要由以下 3 部分组成。

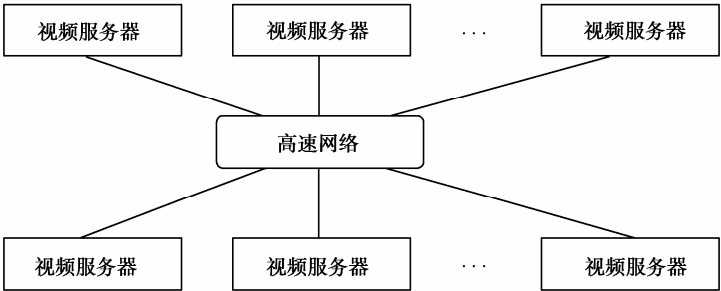


图 10.10 基于 C/S 的 VOD 系统模型

- ① 视频服务器。位于视频点播中心，存储大量的多媒体信息，根据客户的点播请求，把所需的多媒体信息实时地传送给客户。根据系统规模的大小，可采用单一服务器或集群服务器结构来实现。
- ② 高速网络。为视频服务器和客户之间的信息交换提供高带宽、低延迟的网络传输服务。
- ③ 客户端。用户访问视频服务器的工具，可以是机顶盒或计算机，用户通过交互界面将点播请求发送给视频服务器，以及接收和显示来自视频服务器的多媒体信息。

VOD 系统是一种基于客户/服务器模型的点对点实时应用系统，视频服务器可同时为很多用户提供点对点的即时视频点播服务，并且信息交互具有不对称性，客户到视频服务器的上行信道的通信量要远远小于视频服务器到客户的下行信道的通信量。

系统响应时间是 VOD 系统的重要性能指标，它主要取决于视频服务器的吞吐能力和网络带宽。根据系统响应时间长短，VOD 系统可分为真点播 TVOD（True VOD）和准点播 NVOD（Near VOD）两类。

TVOD 要求有严格的即时响应时间，从发出点播请求到接收到节目应小于一秒钟，并提供较完备的交互功能，如对视频的快进、快退和慢放等。TVOD 允许随机地、以任意间隔地对正在播放的视频节目帧进行即时访问，这就对视频服务器的 CPU 处理能力、缓存空间和磁盘 I/O 吞吐量及网络带宽提出很高的要求。

NVOD 对系统响应时间有一定的宽限，从发出点播请求到接收到节目一般在几秒钟到几分钟之间，甚至更长，只要能被用户接受即可。NVOD 将视频节目分成若干时间段而不是帧

进行播放、快进、快退和慢放等操作，时间段比帧的粒度大，从而降低了对系统即时响应的要求，但系统的造价低且支持的客户较多。目前很多 VOD 系统产品都采用 NVOD 方式。

无论是 TVOD 还是 NVOD，当系统规模较大时，单一服务器的处理能力和系统资源都很难满足用户需求，必须通过集群服务器来改进系统性能，提高服务质量。

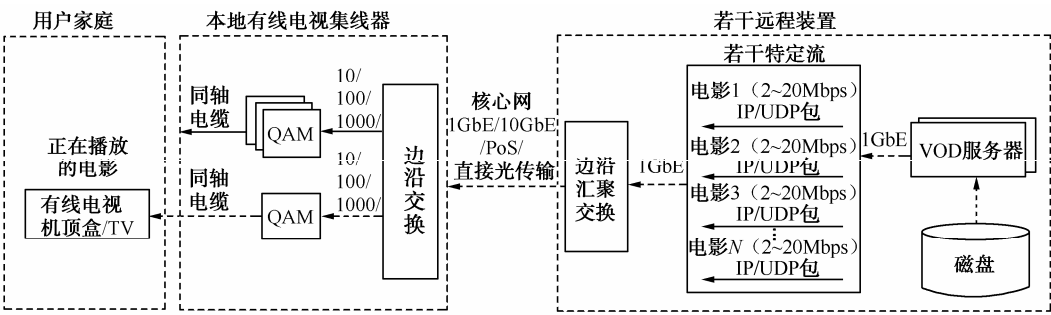
通常一个 VOD 系统可以为用户提供如下视频点播服务。

- ① 影视点播。点播电影或电视节目，用户可以通过快进、快退和慢放等控制功能来控制播放过程。
- ② 信息浏览。浏览各种商品购物和广告信息，或查看股票、证券和房地产行情等信息。
- ③ 远程教育。收看教学节目，选择课程和内容，做练习，模拟考试，自我测试。
- ④ 交互游戏。将视频游戏下载到用户终端上，用户可以和远程的其他用户一起参加游戏。

随着网络环境的改善和 VOD 技术的成熟，VOD 的应用领域将会得到进一步拓展，尤其是在 Internet 上的应用具有广阔的前景。

2. VOD系统组成

VOD 系统主要由显示系统、机顶盒、宽带互动网络系统等组成。如图 10.11 所示是一个简化的 VOD 系统结构图。



注：QAM: Qadrature Amplitude Modulation/正交幅度调制
PoS: Packet over SONET/SDH
1GbE: 千兆位以太网

图 10.11 VOD 系统结构图

(1) VOD 系统的显示系统

VOD 系统的显示系统可由传统的 AV 声像系统及计算机担当。一般来说，欣赏影视片用传统的 AV 声像系统效果较好，查询办公资料用计算机较好、较方便。

(2) VOD 系统的机顶盒

VOD 系统的机顶盒 (Set Top Box, STB) 就是一种数据处理装置，一方面把 VOD 网络上传过来的数字信号转换成传统的 AV 声像系统可播放的多媒体声像信号；另一方面把 VOD 用户的点播指令上传到网络上，指挥信息的播放。普通计算机加装 VOD 专用处理卡及相应的软件，即可起到机顶盒的作用。机顶盒一般要配备遥控器以方便用户使用。

(3) VOD 系统的宽带互动网络系统

VOD 系统的宽带互动网络系统由 VOD 网络、VOD 服务器、VOD 软件组成，起到两个

作用：双向传输多媒体数字信号和点播指令；在服务器端存储及播放多媒体信息。目前，流行的有两大 VOD 网络系统，即有线电视系统和 IP 计算机网络系统。

目前的发展状况是：有线电视系统技术及设备一直不成熟，在试验应用中系统不稳定；功能单一、扩展性较差、升级换代不易；网络与设备复杂，需要对单向有线电视网络进行双向网络改造，造价难以下降，系统用户数量难以达到很大（同时上千户）；没有全球性统一标准，与 Internet、计算机多媒体信息互通与转换复杂，难以做到统一信息平台，也很难跟上计算机网络技术的飞速发展。因此，一直没有较好的应用实例，也难有很好的发展前景。

架构于 IP 计算机网络系统上的 VOD 系统则是最有发展前景的系统，上述有线电视系统的弱点它都不存在，相反，是该系统的优势。VOD 的产生本来就来自于 IP 计算机网络系统。当初人们想用有线电视系统来实现 VOD，是为了借用已有的有线电视系统，即省去对计算机网络的投入，又拥有庞大的现成用户，但在计算机网络投入越来越便宜，Internet 越来越普及，电子商务、家庭办公越来越多地受到人们欢迎的今天，当初采用有线电视系统的理由已不复存在了。相反，IP 计算机网络系统成了酒店、企事业单位、小区一步到位的综合型信息平台，且升级换代极为容易，保护了用户的前期投入。

10.3.4 IP 电话

IP 电话（IP Telephony）、因特网电话（Internet Telephony）和 VoIP（Voice over IP）都是在 IP 网络即分组交换网络上进行的呼叫和通话，而不是在传统的公众交换电话网络（Public Switched Telephone Network, PSTN）上进行的呼叫和通话。当前，IP 电话用于长途通信时的价格比 PSTN 电话的价格便宜得多，但质量也比较低。尽管质量不尽人意，但由于价格上的优势，IP 电话仍然是最近几年来全球多媒体通信中的一个热点技术。

在分组交换网络上传输声音的研究始于 20 世纪 70 年代末和 80 年代初，而真正开发 IP 电话市场始于 1995 年，VocalTec（www.vocaltec.com）公司率先使用 PC 软件在 IP 网络上的两台 PC 之间实现通话。1996 年，科技人员对 IP 网络和 PSTN 网络之间的用户做了第一次通话尝试。1997 年出现了具有电话服务功能的网关，1998 年出现了具有电话会议服务功能的服务器，1999 年是开始应用 IP 电话之年。千禧年开始 IP 电话用在了移动 IP 网络上，如通用分组交换无线服务（General Packet Radio Service, GPRS）或者通用移动电话系统（Universal Mobile Telecommunications System, UMTS）。

IP 电话允许在使用 TCP/IP 协议的因特网、内联网或者专用 LAN 和 WAN 上进行电话交谈。内联网和专用网络可提供比较好的通话质量，与公用交换电话网提供的声音质量可以媲美。在因特网上目前还不能提供像专用网络或者 PSTN 那样的通话质量，但支持保证服务质量（QoS）的协议有望改善这种状况。在因特网上的 IP 电话又叫因特网电话（Internet Telephony），它意味着只要收发双方使用同样的专有软件或者使用与 H.323 标准兼容的软件就可以进行自由通话。通过因特网电话服务提供者（Internet Telephony Service Providers, ITSP），用户可以在 PC 与普通电话（或可视电话）之间通过 IP 网络进行通话。从技术上看，VoIP 比较侧重于声音媒体的压缩编码和网络协议，而 IP 电话比较侧重于各种软件包、工具和服务。

1. IP电话与PSTN电话的技术差别

为了了解 IP 电话和 PSTN 电话在技术上的差别，首先要了解在 IP 网络上拨打 IP 电话和传送声音的过程，如图 10.12 所示。

来自麦克风的聲音在声音输入装置中转换成数字信号，生成“编码声音样本”输出。这些输出样本以帧为单位（如 30 ms 为一帧）组成声音样本块，并复制到缓存存储器。IP 电话应用程序估算样本块的能量。静音检测器根据估算的能量来确定这个样本块是作为“静音样本块”来处理还是作为“说话样本块”来处理。

如果这个样本块是“说话样本块”，就选择一种算法对它进行压缩编码，算法可以是 H.323 中推荐的任何一种声音编码算法或全球数字移动通信系统（Global System for Mobile Communications, GSM）中采用的算法。

在样本块中插入样本块头信息，然后封装到用户数据包协议（UDP）套接接口（Socket Interface）成为信息包。

信息包在物理网络上传送。在通话的另一方接收到信息包之后，去掉样本块头信息，使用与编码算法相反的解码算法重构声音数据，再写入到缓冲存储器。

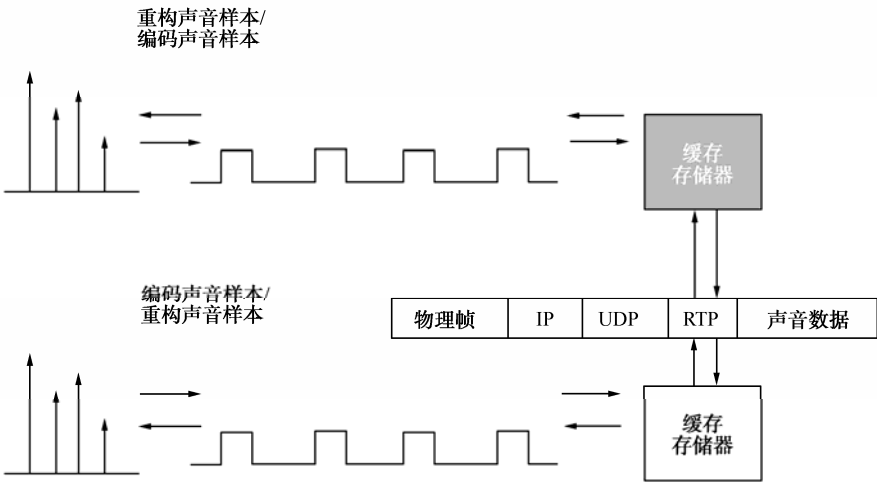


图 10.12 IP 电话的通话过程

从缓存存储器中把声音复制到声音输出设备转换成模拟声音，即可完成一个声音样本块的传送。

从原理上说，IP 电话和 PSTN 电话之间在技术上的主要差别是它们的交换结构。因特网使用的是动态路由技术，而 PSTN 使用的是静态交换技术。PSTN 电话是在电路交换网络上进行的，对每对通话都分配一个固定的带宽，因此通话质量有保证。在使用 PSTN 电话时，呼叫方拿起收/发话器，拨打被呼叫方的国家码、地区码和市区号码，通过中央局建立连接，然后双方就可进行通话了。在使用 IP 电话时，用户输入的电话号码转发到位于专用小型交换机（Private Branch Exchange, PBX）和 TCP/IP 网络之间最近的 IP 电话网关，IP 电话网关查找通过因特网到达被呼叫号码的路径，然后建立呼叫。IP 电话网关把声音数据装配成 IP 信息包，然后按照 TCP/IP 网络上查找到的路径把 IP 信息包发送出去。对方的 IP 电话网关接收到这种 IP 信息包之后，把信息包还原成原来的声音数据，并通过 PBX 转发给被呼叫方。

2. IP电话的通话方式

IP 电话有 3 种基本的通话方式：在 IP 终端（计算机）之间的通话、IP 终端与普通电话（或可视电话）之间通过 IP 网络和 PSTN 网络的通话、普通电话（或可视电话）之间通过 IP 网络和 PSTN 网络的通话。

IP 终端之间的通话方式如图 10.13 所示。在这种通话方式中，通话收发双方都要使用配置了相同类型的或者兼容的 IP 电话软件和相关部件，如声卡、麦克风、喇叭等。声音的压缩和解压缩由 PC 承担。



图 10.13 IP 终端之间的通话方式

IP 终端与电话终端之间的通话方式如图 10.14 所示。在这种通话方式中，通话的一方使用配置了 IP 电话软件和相关部件的计算机，另一方则使用 PSTN/ISDN/GSM 网络上的电话。在 IP 网络的边沿需要有一台配有 IP 电话交换功能的网关，用来控制信息的传输，并且把 IP 信息包转换成电路交换网络上传送的声音，或者相反。



图 10.14 IP 终端与电话终端之间的通话方式

通过 IP 网络的电话之间的通话方式如图 10.15 所示。在这种方式中，通话双方都使用普通电话或一方使用可视电话，或者双方都使用可视电话。这种方式主要用在长途通信中，在通话双方的 IP 网络边沿都需要配置具有电话功能的网关，以便进行 IP 信息包和声音之间的转换及控制信息的传输。

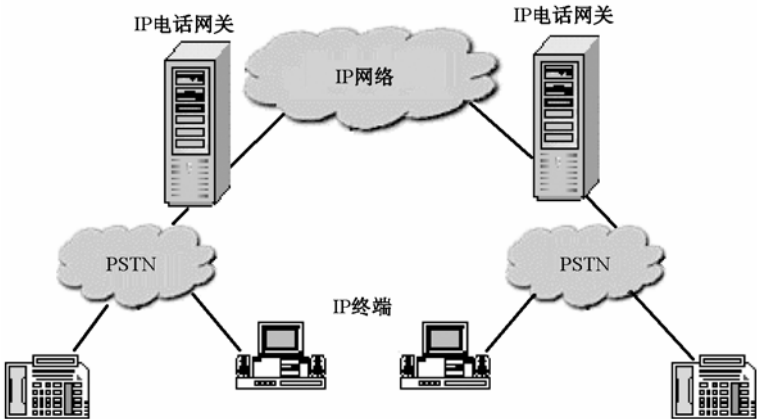


图 10.15 通过 IP 网络的电话之间的通话

3. IP电话标准

开通 IP 电话服务使用的一个重要标准是信号传输协议 (Signalling Protocol)。信号传输协议是建立和控制多媒体会话或者呼叫的一种协议, 数据传输 (Data Transmission) 不属于信号传输协议的内容。这些多媒体会话包括多媒体会议、电话、远距离学习和类似的应用。IP 信号传输协议 (IP Signalling Protocol) 用来创建网络上客户的软件和硬件之间的连接。多媒体会话的呼叫建立和控制功能包括用户地址查找、地址转换、连接建立、服务特性磋商、呼叫终止和呼叫参与者的管理等。附加的信号传输协议包括账单管理、安全管理、目录服务等。

广泛使用 IP 电话的最关键问题之一是建立国际标准, 这样可使不同厂商开发和生产的设备能够正常地在一起工作。当前开发 IP 电话标准的组织主要有 ITU—T, IETF 和欧洲电信标准学会 (European Telecommunications Standards Institute, ETSI) 等。人们认为两个比较值得注意的可用于 IP 电话信号传输的标准是 ITU 的 H.323 系列标准和 IETF 的入会协议 (Session Initiation Protocol, SIP)。SIP 是由 IETF 的 MMUSIC (Multiparty Multimedia Session Control) 工作组正在开发的协议, 它是在 HTML 语言基础上开发的、比 H.323 简便的一种协议。该协议原来是为在因特网上召开多媒体会议开发的协议。H.323 和 SIP 这两种协议代表解决相同问题 (多媒体会议的信号传输和控制) 的两种不同的解决方法。此外, 还有两个信号传输协议被考虑为 SIP 结构的一部分。这两个协议是会话说明协议 (Session Description Protocol, SDP) 和会话通告协议 (Session Announcement Protocol, SAP)。iNOW!、因特网多媒体远程会议协会 (International Multimedia Teleconferencing Consortium, IMTC) 的 VoIP forum 和 MIT 因特网电话协会 (MIT Internet Telephony Consortium) 对不同标准和网络之间的协同工作比较感兴趣。

习 题 10

1. 填空题

- (1) 超文本是指_____。
- (2) 超媒体是指_____。
- (3) HTML 的意思是_____。
- (4) 多媒体通信 (Multimedia Communications) 是_____与_____的完美结合。
- (5) _____中文名称是综合业务数字网, 俗称“一线通”。
- (6) 根据系统响应时间长短, VOD 系统可分为_____和_____两类。
- (7) 目前, 流行的有两大 VOD 网络系统, 即_____和_____。

2. 简答题

- (1) 多媒体通信与传统的通信方式相比有哪些特点?
- (2) 多媒体通信需要解决哪些关键技术?
- (3) 传统的通信网络可以分为哪些类型?
- (4) 分别以 B-ISDN, FDDI, Ethernet, ATM 和 VOD 为例, 概述多媒体信息传输的特点。
- (5) 什么是视频会议? 什么是 IP 电话? 试简述它们的相关标准的具体内容。

(6) 多媒体通信的标准有哪些?

3. 操作题

选择一个网页制作工具 (FrontPage 或 Dreamweaver), 设计一个简单的 HTML 网页, 介绍自己的基本情况。

第二部分 操作实验

实验1 数字声音

实验目的

- (1) 了解音频处理软件的相关知识，掌握音频处理的基本方法；
- (2) 了解音频处理软件 CoolEdit 的主要功能；
- (3) 掌握通过 Windows 系统自带的录音软件获取声音的基本方法；
- (4) 掌握通过 CD 光盘获取声音的基本方法；
- (5) 掌握 CoolEdit 软件降低噪声的基本方法；
- (6) 掌握 CoolEdit 软件添加音效的基本方法。

实验内容

T1.1 Windows录音机获取声音实例

在 Windows 操作系统中自带的“录音机”程序可以简单处理声音。选择“开始”→“程序”→“附件”→“娱乐”→“录音机”之后即可调出处理声音的窗口。使用它可录音，进行简单的声音编辑（如插入、删除等）。

如果你的计算机安装有声卡，一般来说都附带有声音工具。例如，声霸（Sound Blaster）卡带有几种声音工具，通常要由用户自己安装。其中，功能比较强的是 Creative WaveStudio。它是一种方便易用的音频编辑工具，可播放、编辑和录音 8 位（磁带质量）、16 位（CD 质量）和 24 位的音频数据，使用各种特殊效果和编辑操作，如采用重复、反转、回声、位移、剪切、复制和粘贴来增强你的音频数据或建立独特的声音，可支持 DirectX 音频插件，可以同时编辑多个音频文件，支持 RAW，MP3 和 Windows Media Audio（WMA）等格式的音频文件。

1. 安装声卡、连接话筒和音箱

在录制之前，先将麦克风连接到声卡上。一般声卡在主机机箱后有 3 个插孔，输出（Speaker Out）、线性输入（Line In）和麦克风（Microphone）。其中，输出和耳机（音箱）插头相连，话筒插头自然应该插入麦克风插孔，如图 T1.1 所示。

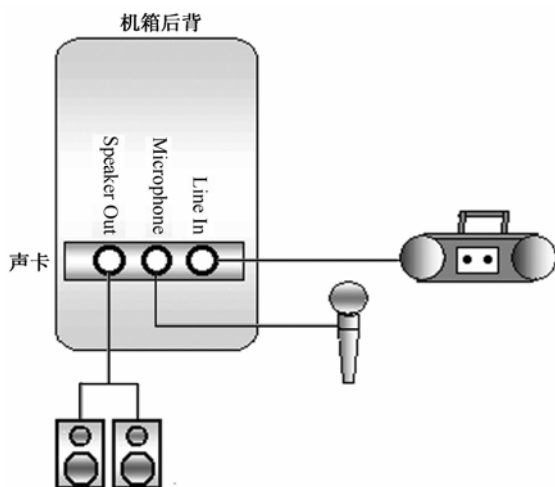


图 T1.1 声卡与其他设备相连

2. 在Windows中进行录音设置

在 Windows 系统默认情况下，麦克风的音量开关是关闭的，即使将麦克风插好了也不能将麦克风声音录下来，要在 Windows 中打开麦克风的录音开关。具体操作如下：

(1) 双击 Windows 任务栏右端的“音量”小喇叭，弹出的面板如图 T1.2 所示。

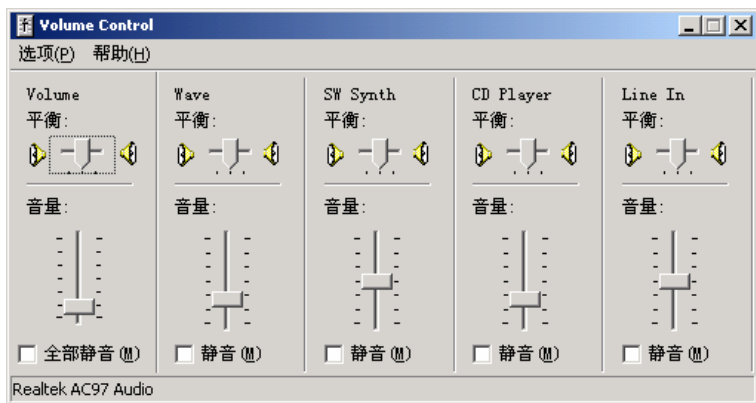


图 T1.2 音量控制面板

(2) 选择“选项”→“属性”菜单，在弹出的对话框中选择“录音”项，在“显示下列音量控制”滚动列表框中，选中“Microphone”复选框，如图 T1.3 所示。单击“确定”按钮后弹出录音控制面板（Recording Control），如图 T1.4 所示。选择“Microphone”作为录音设备。

(3) 取消麦克风的播放静音状态（可选步骤）。

如果希望在录音的同时，耳机能听到麦克风输入的声音，必须取消麦克风的播放静音状态。

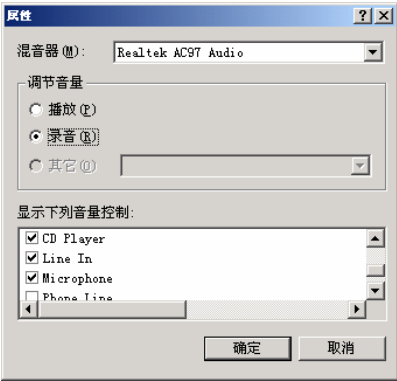


图 T1.3 “属性”对话框



图 T1.4 录音控制面板

重回“属性”对话框，选择“播放”项，选中“显示下列音量控制”中的“Microphone”，如图 T1.5 所示。单击“确定”按钮后弹出“Volume Control”面板（音量控制面板）。

在“Volume Control”面板中取消“Microphone”的静音状态，如图 T1.6 所示。



图 T1.5 “属性”对话框

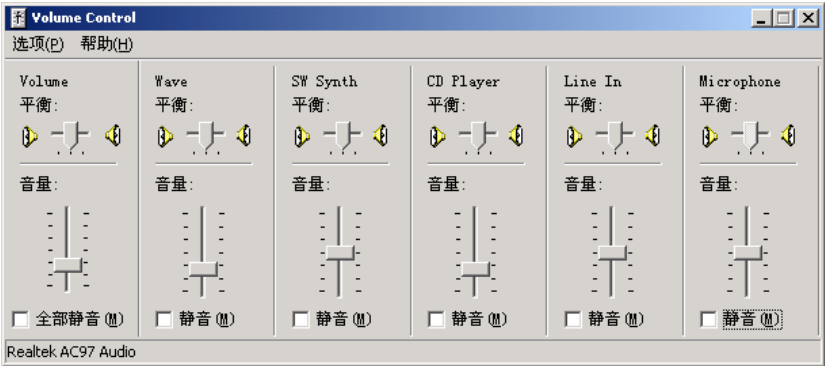


图 T1.6 音量控制面板

3. “录音机” 录音




- (1) 单击“开始”按钮，选择“更多程序”→“附件”→“娱乐”→“录音机”命令，打开“声音-录音机”窗口，如图 T1.7 所示。
- (2) 单击录音按钮，即可开始录音，录音长度最长为 60s。如果设置正确，在窗口中间的波形显示窗内可以看见有波形图随声音的变化而变化，如图 T1.8 所示。



图 T1.7 “声音-录音机”窗口



图 T1.8 录音机录音状态

- (3) 录制完毕后，单击停止按钮即可。
- (4) 单击播放按钮，即可播放所录制的声音文件。
- (5) 保存音频文件。选择菜单“文件”→“保存”，所录制的声音以波形（.wav）文件保存。

T1.2 Easy CD-DA Extractor获取CD声音实例

对一名音乐爱好者来说，一定会拥有很多的 CD，从 CD 中抓取音乐精品也就是抓轨的过程，Easy CD-DA Extractor（Version 10b）是一款优秀的音乐 CD 抓取、格式转换、光盘刻录软件。

1. 安装Easy CD-DA Extractor软件

安装方法略。

2. 获取CD光盘声音



- (1) 准备一张歌曲或音乐 CD 光盘，插入光盘驱动器。
- (2) 启动 Easy CD-DA Extractor 软件，显示主界面，如图 T1.9 所示。在主界面中自动列出光盘中的曲目清单，该曲目清单又叫“音轨”，是声音在光盘上的存储单元。
- (3) 选择要获取的曲目。方法是：单击某个音轨行，该行显示黑色光标条，然后可在界面左下部选择“CD 播放器”标签，单击播放按钮，试听该曲目。确认后，单击停止按钮，如图 T1.10 所示。



图 T1.9 “Easy CD-DA Extractor”主界面

(4) 单击已经确认的音轨左端的方框，使其中显示“√”，如图 T1.11 所示。

(5) 参数设置。单击左下部的“输出”标签，回到主界面，如图 T1.9 所示。在设置输出文件格式下拉列表中选择需要的格式（支持 MP3, WMV, AAC, VQF, WAV, AIFF 等格式）。在设置采样频率下拉列表中选择合适的采样频率（支持接近 CD 品质、CD 品质、320kbps、256kbps 等采样频率）。在设置文件输出路径文本框中指定一个保存音频文件的文件夹，比如“C:\My Music”。在输出文件名称框中输入文件名格式，如“<SHE> - <Super Star>\<编号>.<标题>”。



图 T1.10 Easy CD-DA Extractor 中的 CD 播放器

所有曲目相同艺人					
	#	艺人	标题	开始	长度
<input checked="" type="checkbox"/>	01		曲目 1	00:02.00	03:16.73
<input type="checkbox"/>	02		曲目 2	03:18.73	04:10.05
<input checked="" type="checkbox"/>	03		曲目 3	07:29.03	05:14.43
<input type="checkbox"/>	04		曲目 4	12:43.46	03:42.10
<input checked="" type="checkbox"/>	05		曲目 5	16:25.56	03:56.01
	19				78:02.15

图 T1.11 被选中的音轨

(6) 单击“转录 CD”按钮，开始获取音频文件。在获取过程中，会显示如图 T1.12 (a)

所示的获取进程窗口。当获取进程结束后，获取进程画面底部的按钮由“取消”变成“关闭”，如图 T1.12（b）所示，单击该按钮，结束操作。



图 T1.12 获取进程窗口

（7）观察“C:\My Music\ SHE - Super Star”文件夹，其中应该有“01. 曲目 1.mp3”、“03. 曲目 3.mp3”、“05. 曲目 5.mp3”文件，打开上述文件，聆听声音，检查声音抓取效果。

T1.3 CoolEdit Pro制作个人音乐实例

用 Windows 系统自带的录音机软件，可以做一些简单的录音、剪切、混合等操作，但这远远不能满足需要。因特网上许多站点提供试用的或者免费的声音工具。CoolEdit Pro 是一个功能强大的音乐编辑软件，最高采样频率可以达到 192kHz，最高量化精度可以达到 32 位，支持 22 种音乐文件格式，能够高质量地完成录音、编辑、合成等多种任务，其完整的声音与音效的处理为用户提供了完整的音乐解决方案。只要拥有它和一台配备了声卡的计算机，就等于同时拥有了一台多轨数码录音机、一台音乐编辑机和一台专业合成器。

CoolEdit Pro 不仅适合于专业人员，也适合普通音乐爱好者。类似的工具还有 GoldWave，CakeWalk，Cubase 等。下面就用 CoolEdit Pro 2.0 来制作个人音乐。

1. 录制原声

录音是所有后期制作加工的基础，这个环节出问题，是无法靠后期加工来补救的，所以，如果是原始的录音有较大问题，就重新录吧。录音的环境设置参见 T1.1 节。

安装好 CoolEdit Pro 2.0, 打开 CoolEdit Pro 2.0, 会自动建立一个新的工程，界面如图 T1.13 所示。



图 T1.13 CoolEdit Pro 软件主界面

(1) 单击多轨与单轨切换的按钮,如图 T1.13 所示,进入多音轨界面,右键单击音轨 1 空白处,插入你所要录制歌曲的 MP3 (WAV 也可) 伴奏文件,如图 T1.14 所示。


(2) 选择将你的声音录在音轨 2，按下 按钮，如图 T1.15 所示。



图 T1.14 插入伴奏文件



图 T1.15 选择录音音轨

(3) 按下左下方的红色录音键，跟随伴奏音乐开始演唱和录制，如图 T1.16 所示。

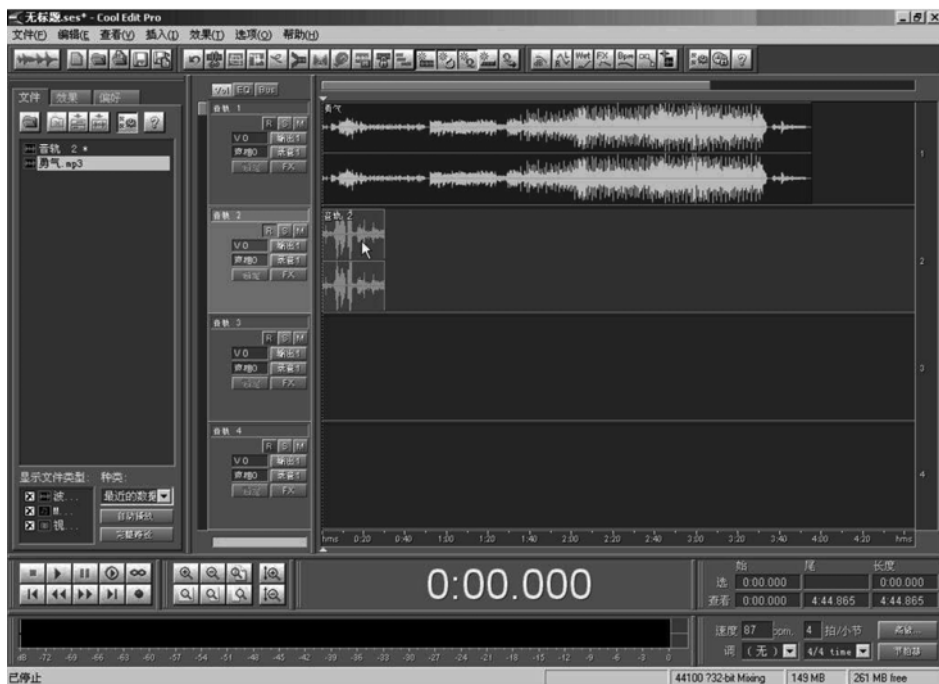


图 T1.16 演唱和录制界面


(4) 录音完毕后（再次单击红色录音键，停止录音），可单击左下方播放按钮进行试听，看有无严重的出错，是否要重新录制，如图 T1.17 所示。



图 T1.17 试听录音

(5) 双击音轨 2 进入波形编辑界面，如图 T1.18 所示。



图 T1.18 波形编辑界面

(6) 选择菜单“文件”→“另存为”，将你录制的自己的原声文件保存为 MP3 格式，如图 T1.19 所示。



图 T1.19 保存录制的原声

需要说明的是：录制时要关闭音箱，通过耳机来听伴奏，跟着伴奏进行演唱和录音。另外，如果你的麦克风总是录入从耳机中传出的伴奏音乐的声音，建议你用普通的大话筒，只要加一个大转小的接头即可直接在计算机上使用，你会发现录出的效果要干净得多。

2. 降噪处理

降噪是至关重要的一步，做得好有利于下面进一步美化你的声音，做不好就会导致声音失真，彻底破坏原声。


(1) 单击多轨与单轨切换的按钮（如图 T1.13 所示），回到多轨模式下，在第 3 轨处按下 **R** 按钮，单击录音键 ，不要出声，录下一段空白的噪声文件，不需要很长，如图 T1.20 所示。



图 T1.20 录制环境噪声

(2) 当录制完成后，双击该波形进入单轨模式，选择菜单“效果” → “噪音消除” →

“降噪器”，单击“噪音采样”按钮，然后单击“关闭”按钮（注意，这里只单击“关闭”按钮，不单击“确定”按钮），如图 T1.21 所示。

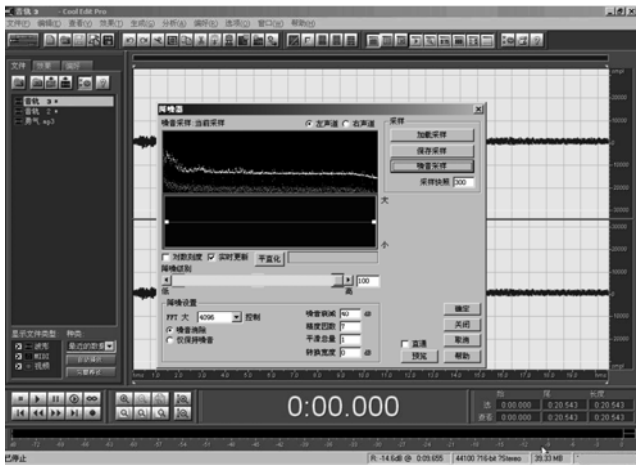


图 T1.21 对环境噪声进行采样

(3) 单击多轨与单轨切换的按钮（如图 T1.13 所示），回到多轨模式下，右键单击刚才录制的环境噪声波形，从弹出的快捷菜单中选择“清除音块（移除/关闭）”，删除此轨的波形（不需要保留）。

(4) 双击录制声音所在的轨道，进入单轨模式，单击“效果”菜单中的“降噪器”，我们在前面已经进行过了环境的噪声采样，此时只需单击“确定”按钮，降噪器就会自动消除录制声音中的环境噪声，如图 T1.22 所示。



图 T1.22 消除录制原声中的环境噪声

3. 声效处理

为了让声音变得更好听，需要进行效果处理。以下的效果处理需要安装 BBE, Ultrafunk2, Waves 插件。安装完毕后，选择“效果”菜单下的“刷新效果列表”刷新一下 CoolEdit Pro 2.0

的效果列表，你会发现 DirectX 下多了很多的选项，这就是安装的插件，如图 T1.23 所示。

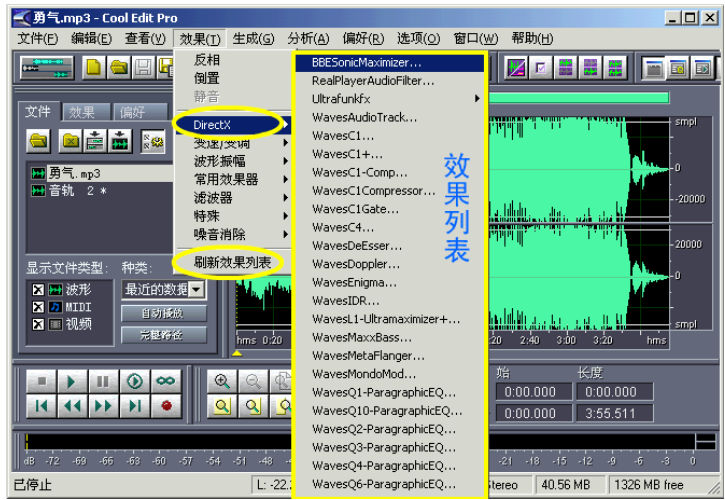


图 T1.23 音效插件菜单

1) 高音激励处理

选择菜单“效果” → “DirectX” → “BBE Sonic Maximizer”，打开 BBE 高音激励器，如图 T1.24 所示。加载“软件预置”下拉菜单中的各种效果（或全手动调节三旋钮）后，单击激励器右下方的“预览”按钮进行反复的试听，直至调至满意的效果后，单击“确定”按钮，对原声进行高音激励。

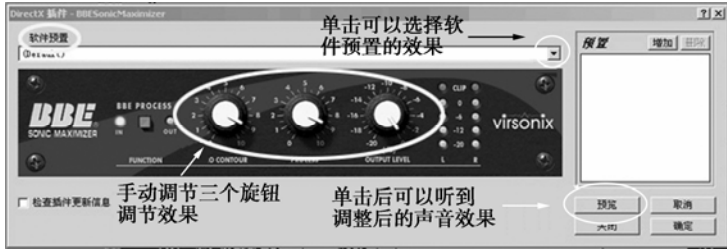


图 T1.24 BBE 高音激励器

注：此过程的目的是为了调节所录人声的高音和低音部分，使声音显得更加清晰明亮或厚重。激励的作用就是产生谐波，对声音进行修饰和美化，产生悦耳的听觉效果，它可以增强声音的频率动态，提高清晰度、亮度、音量、温暖感和厚重感，使声音更有张力。

2) 压限处理

选择菜单“效果” → “DirectX” → “WavesC4...”，打开 WavesC4 压限效果器，如图 T1.25 所示。加载“软件预置”下拉菜单中的各种效果中的某一种或单击 **Load** 按钮，从弹出的列表中进行选择（如果你对数字音频有足够了解的话，也可手动调节）。单击右下方的“预览”按钮进行反复的试听，直至调至满意的效果后，单击“确定”按钮，对原声进行压限处理（对人声的处理“Pop vocal”效果比较好）。

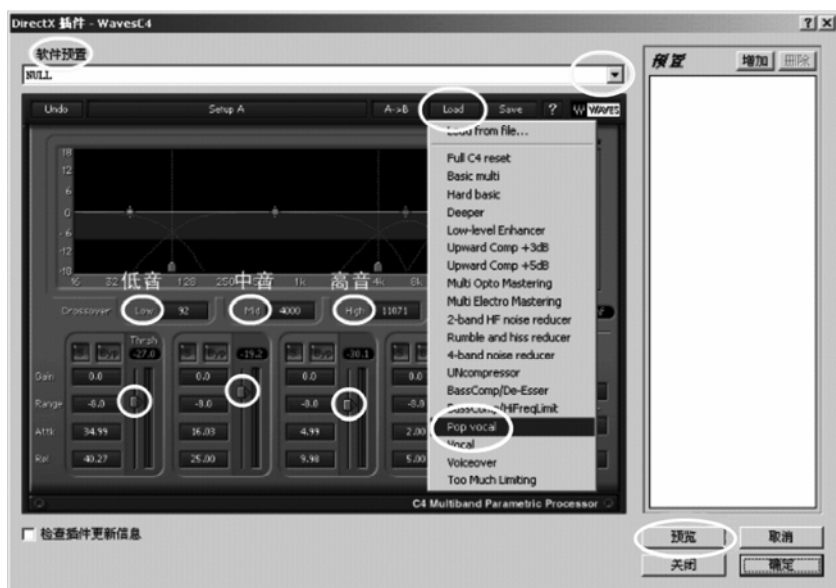


图 T1.25 WavesC4 压限效果器

注：压限的目的就是把你的声音通过处理后变的更加均衡，保持一致连贯，声音不至忽大忽小，忽高忽低。

3) 混响处理

选择菜单“效果”→“DirectX”→“Ultrafunkfx”→“Reverb…”，打开混响效果器，如图 T1.26 所示。加载“软件预置”下拉菜单中的各种效果后（也可手动调节），单击右下方的“预览”按钮进行反复的试听，直至调至满意的混响效果后，单击“确定”按钮，对原声进行混响处理。

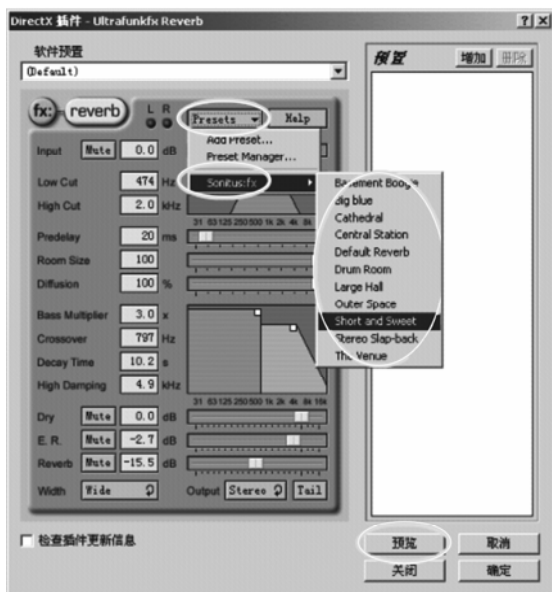


图 T1.26 混响效果器

注：在混响处理后，可以使你的声音显得不那么干涩，变的圆润和厚重一些。注意，如果混响加得太多的话会使人声显得模糊过于假，加得少了又会使声音显得很干很涩，可以慢慢摸索，多录几次后就有经验了。

4. 混缩合成

(1) 选择菜单“编辑”→“混缩到文件”→“全部波形”，便可将伴奏和处理过的人声混缩合在一起，如图 T1.27 所示。

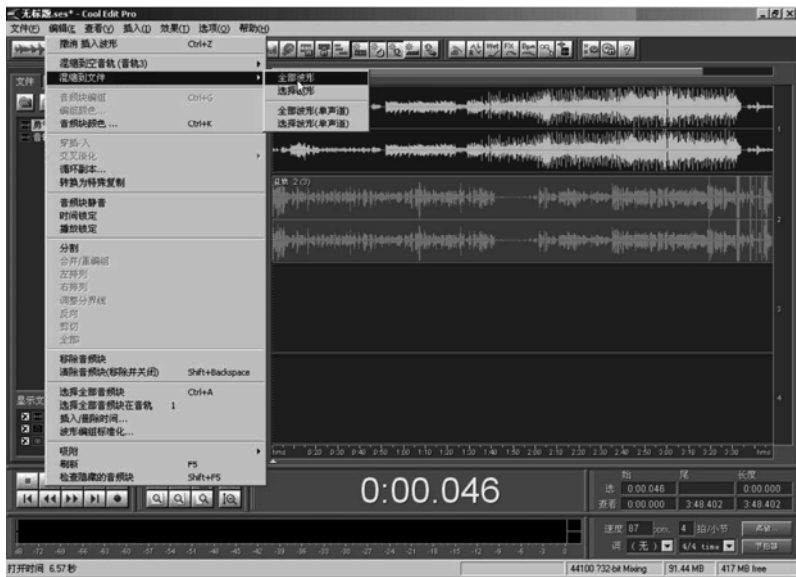


图 T1.27 混缩到文件

(2) 选择菜单“文件”→“另存为”，将混缩合成后的文件存为 MP3 格式，如图 T1.28 所示。

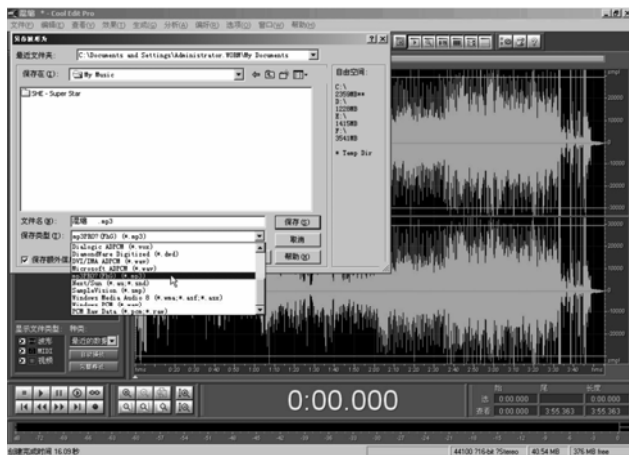


图 T1.28 保存混缩合成文件

至此，个人音乐就制作完成了，如果有时间你还可以给自己的歌加上一些自己摄制的外景、配上字幕，做成 VCD 或 DVD 光盘，这样一张个人音乐专辑就诞生了。

实验思考

- (1) 数字音频通常用的采样频率是多少?
- (2) 多媒体音频技术中常用的存储声音信息的文件格式有哪些?
- (3) 音频文件的数据量与哪些因素有关?
- (4) 怎样通过 CD 光盘获取数字音频?
- (5) 如果要录制小于 1 分钟的声音, 可以使用什么方法?
- (6) 如果要录制任意时间长度的声音, 使用什么方法?
- (7) 在录制声音时, 如果发生录不到声音的现象, 应该如何解决?
- (8) 音频处理软件 CoolEdit 的主要功能是什么?
- (9) 如何降低噪声?
- (10) 如何添加音效?
- (11) 如何对声音进行混缩合成?

实验2 数字图像

实验目的

- (1) 了解数字图像的技术参数及文件格式;
- (2) 掌握 Windows 界面获取的基本方法;
- (3) 掌握对图像进行简单加工的基本方法;
- (4) 掌握对图像进行浏览、移动、复制、文件格式转化的基本方法;
- (5) 掌握图像处理软件 Photoshop 的一些基本图像处理方法。

实验内容

T2.1 Windows 屏幕图像获取、加工和使用实例

1. Windows 界面获取

Windows 界面包括在 Windows 环境下运行的各类软件的界面、窗口、菜单等显示信息，在多媒体产品的使用说明、操作步骤、产品介绍等场合应用非常广泛。

获取界面的操作习惯上又叫“插图”，可以直接在键盘上进行，也可以借助专用软件进行。在键盘上进行时，有些界面需要进行图像处理，如菜单界面。

1) 获取整个屏幕图像

获取整个屏幕的操作步骤如下：

- (1) 在屏幕上打开若干个软件的界面，或者显示菜单、文字、图片等其他信息。
- (2) 按 **Print Screen** 键，整个屏幕的显示信息被保存在 Windows 的剪贴板中。
- (3) 打开 Word，新建文档，选择“编辑”→“粘贴”菜单命令，整个屏幕图像就会出现在新建文档中。

如果将获取整个屏幕的图像作为图像文件保存，则继续以下步骤。

(4) 打开 Windows 系统的“画图”软件，选择“编辑”→“粘贴”菜单命令，把 Windows 剪贴板中的屏幕显示信息粘贴到画图窗口中，从而获得界面图像。可以用“画图”软件工具对图进行修改，最后以 **BMP** 或 **JPEG** 格式保存到文件中。

2) 获取屏幕当前窗口图像

获取当前窗口的操作步骤如下：

- (1) 在屏幕上经常显示很多窗口，单击要获取的窗口，将其变成当前活动窗口。
- (2) 按下 **Alt+Print Screen** 组合键，当前窗口被复制到 Windows 的剪贴板中。
- (3) 打开 Windows 系统的“画图”软件，选择“编辑”→“粘贴”菜单命令，把保存在 Windows 剪贴板中的屏幕显示信息粘贴到窗口中，从而获得当前窗口图像。

说明：如果想获取带菜单显示的当前窗口，如图 T2.1 所示，不能用鼠标选择菜单，然后再按下 Alt+Print Screen 组合键来获取当前窗口，因为 Alt 键是热键，一旦按下，菜单会消失。正确的做法是用快捷键打开菜单，如 Alt+F 打开“文件”菜单，然后再按下 Print Screen 键。

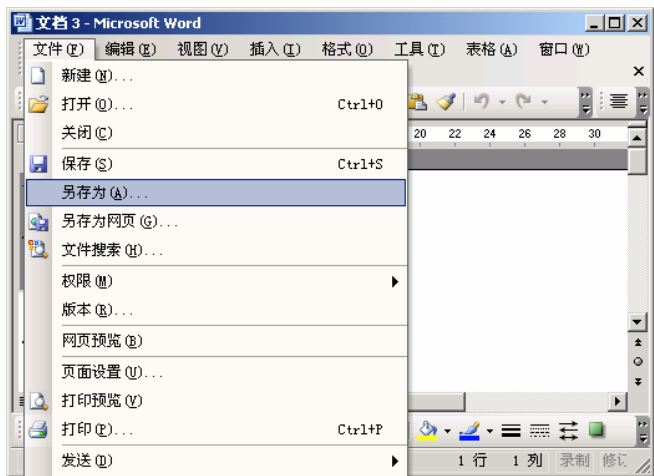


图 T2.1 带菜单显示的当前窗口

2. 对图像文件进行简单加工

可以使用 Windows 系统的“画图”软件对图像进行简单的加工处理，操作步骤如下：
(1) 打开 Windows 系统的“画图”软件，如图 T2.2 所示。

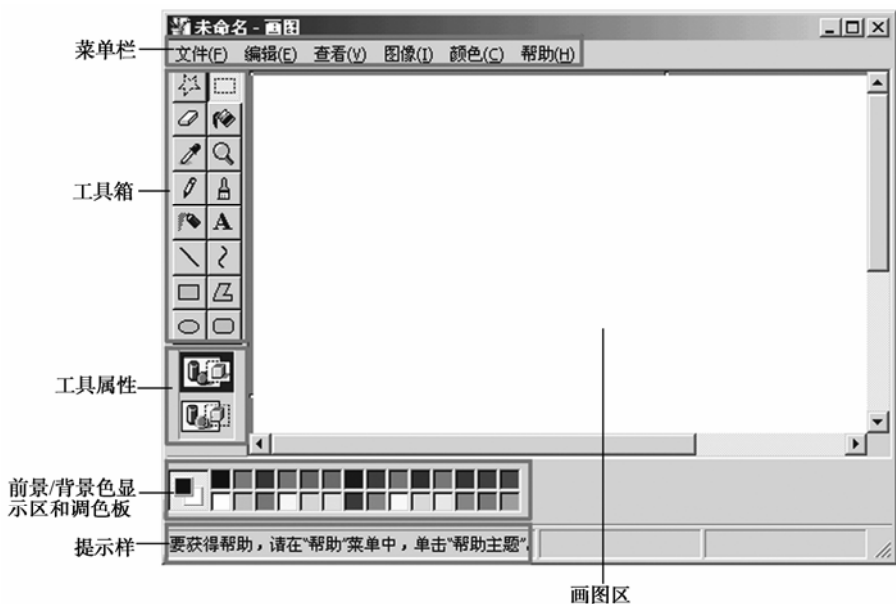


图 T2.2 Windows 系统的“画图”软件窗口

(2) 打开图像文件。在“画图”窗口中选择“文件” → “打开”菜单命令，在弹出的

“打开”对话框中选择打开一个“BMP”（或“JPEG”）文件。

（3）改变图像尺寸。在“画图”窗口中选择“图像”→“拉伸/扭曲”菜单命令，显示“拉伸和扭曲”对话框，如图 T2.3 所示，在“水平”和“垂直”两个输入框中分别输入数值，如 50，意思是把图像缩小到原图的 50%。

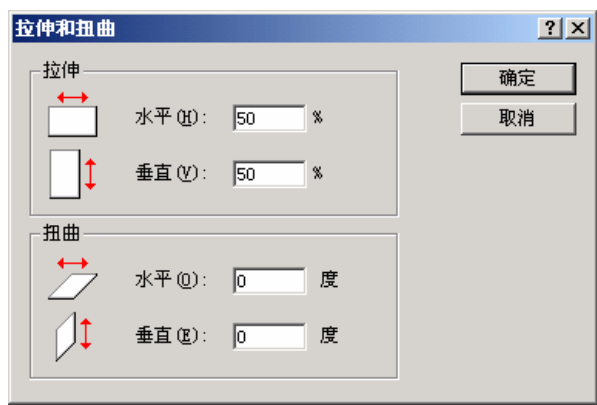


图 T2.3 “拉伸和扭曲”对话框

（4）添加文字。在“画图”窗口中单击“文字”按钮，用鼠标左键在图像上拖出一个文字书写区域，然后添加文字。文字背景如果带有底色，单击“透明”按钮即可消除，如图 T2.4 所示。

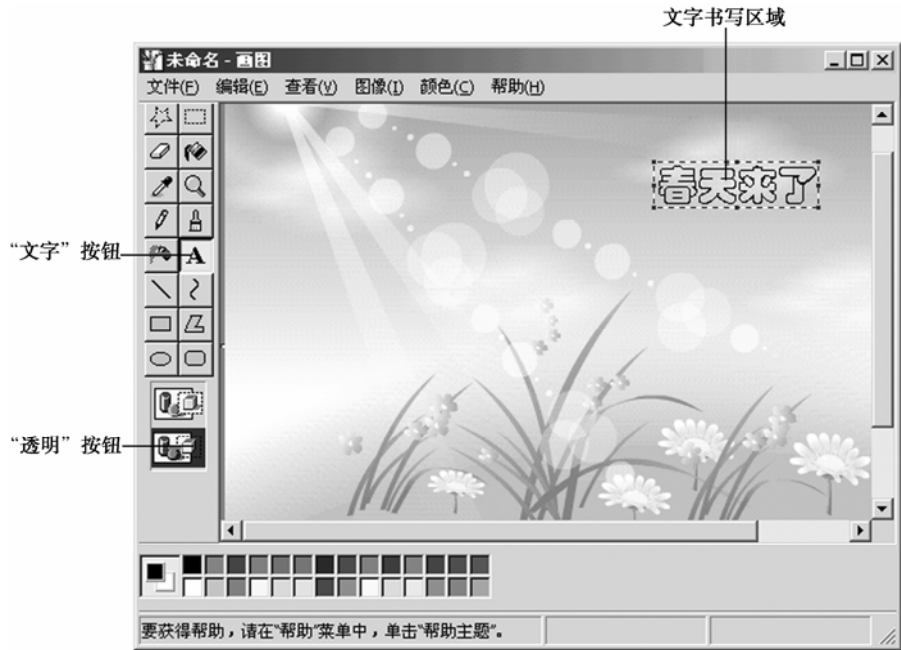


图 T2.4 “文字”按钮、“透明”按钮和“文字书写区域”的位置

把鼠标对准并拖动文字输入框，即可移动文字的位置。最后单击文字输入框以外的区域来结束文字操作。

(5) 保存上面的文件。在“画图”窗口中选择“文件”→“保存”菜单命令，在弹出的“保存为”对话框（如图 T2.5 所示）中依次进行如下操作：

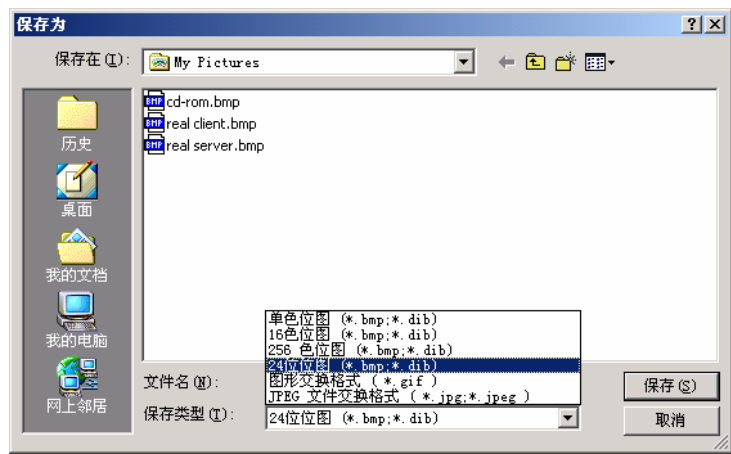


图 T2.5 “保存为”对话框

首先，选择“24 位位图”类型，文件取名为“type-24b”，单击“保存”按钮。
其次，选择“256 色位图”类型，文件取名为“type-256b”，单击“保存”按钮。
最后，选择“16 色位图”类型，文件取名为“type-16”，单击“保存”按钮。
当选择“256 色位图”和“16 色位图”类型保存时，显示如图 T2.6 所示的对话框，单击“是”按钮。

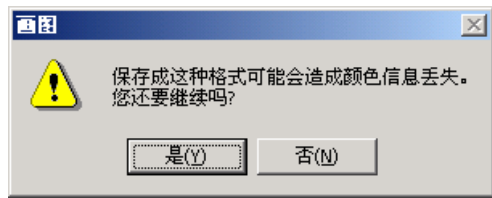


图 T2.6 “画图”提示信息对话框

对比保存的 3 种类型图像的差异，即可理解颜色深度对图像质量的影响。

T2.2 Photoshop图像处理实例

图像处理软件非常多，常见的有 Photoshop，PhotoStyler，PCPaintBrush，CorelDraw 等。其中 Photoshop 以其直观的界面、全面的功能成为最流行的图像处理软件，是我们学习的首选软件。

Photoshop 是由美国 Adobe 公司开发的专用图像处理软件，集图像创作、扫描、编辑、修改、合成于一体。Adobe 公司于 1990 年推出 Photoshop 软件，2003 年推出的 Photoshop CS 版，功能更加强大。

1. 初识Photoshop CS

Photoshop CS 的操作界面主要分为菜单栏、工具属性栏、工具箱、控制面板及图像编辑

窗口这几部分，如图 T2.7 所示。



图 T2.7 Photoshop CS 界面

- (1) 菜单栏中有一系列的命令，在它上面几乎囊括了所有 Photoshop 对图像编辑、处理的操作命令。
- (2) 工具属性栏的用途是提供当前所用工具的有关信息、可以使用的功能及设置参数等。
- (3) 工具箱包含了 Photoshop CS 所有的使用工具，部分工具按钮右下角有一个小三角，说明它是一个工具组，用鼠标单击小三角就会出现其他工具按钮，如图 T2.8 所示。

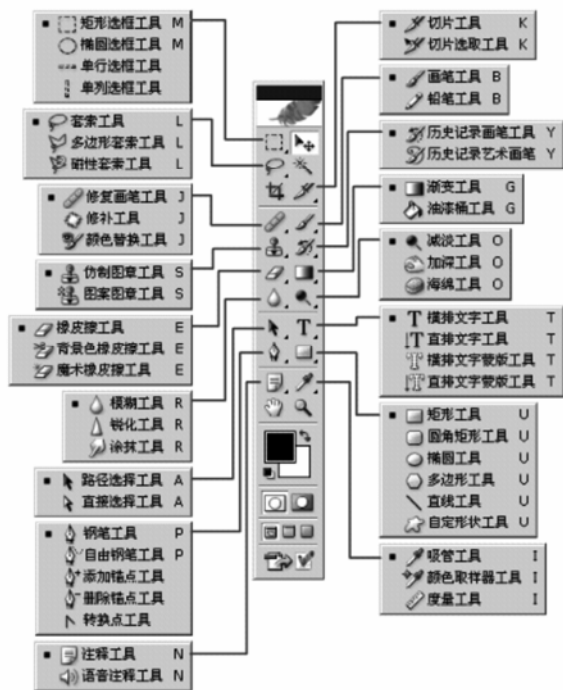


图 T2.8 工具箱

(4) 控制面板位于编辑窗口的右侧，在默认的情况下分为 3 部分：导航器和信息控制板；颜色、色板和样式控制板；图层、通道、路径、历史记录和动作控制板。每一个控制板上都有一些标签，单击某个标签便可显示相应的控制面板，并且显示有关图像的信息。利用菜单栏的“窗口”命令可以打开或关闭控制面板。

(5) 图像编辑窗口是 Photoshop 的工作区，用来显示图像文件，进行浏览和编辑。

2. Photoshop 的基本概念

1) Photoshop 的图层

打开 Photoshop 自带的“鲜花.psd”文件（在 Photoshop 安装目录的“样本”子目录下），其效果及图层结构如图 T2.9 所示。



图 T2.9 Photoshop 作品效果及图层结构

效果图是由各个图层拼合而成的，上面的图层覆盖下面的图层（向日葵图层覆盖了花卉背景图层）。图层还有着更多的功能、特效和使用方法，可以自由地设置透明度，调整次序，可以拿出来单独进行修改。在图层调板的某一层上单击，该层就成为当前层，可以进行编辑修改和添加特效等操作。

利用图层的这些特性，可以巧妙地将不同的图像进行组合，有时几张很平常的图像，就能组合成非常具有艺术感染力的作品。

2) Photoshop 的通道

同图层一样，通道实质上也是将图像分成独立的几个部分，不过分割的标准不是层次关系而是不同的色彩。

可以把通道看做某一种色彩的集合，如红色通道，记录的就是图像中不同位置红色的深浅（即红色的灰度），除了红色外，在该通道中不记录其他颜色的信息。大家知道，绝大部分的可见光可以用红、绿、蓝三原色按不同的比例和强度混合来表示，将三原色的灰度分别用一个颜色通道来记录，最后合成各种不同的颜色，计算机显示器使用的就是 RGB 模型，Photoshop 中默认的颜色模式也是 RGB。

在图层调板中单击“通道”选项卡，就可看到通道调板，如图 T2.10 所示，单击 R，G，B 中的一个通道将关闭其他通道。

以上所说的是 RGB 颜色模式，在 Photoshop 中还有 CMYK，HSB，Lab 等颜色模式，不同的模式有不同的默认颜色通道数目和类型，每一模式下通道分解和合成的基本原理是一致的。

3) Photoshop 的滤镜

滤镜 (Filter) 是 Photoshop 的特色工具之一，充分而适度地利用好滤镜不仅可以改善图像效果、掩盖缺陷，还可以在原有图像的基础上产生许多特殊的效果。

滤镜是一种植入 Photoshop 的外挂功能模块，世界上有很多公司开发了各种各样的滤镜插件来制作特效。Photoshop 滤镜的安装很简单，一般滤镜文件的扩展名为 .8BF，只要将这个文件复制到 Photoshop 安装目录下的“滤镜”目录下就可以了。Photoshop 还自带了许多好用的滤镜。如图 T2.11 所示就是用 Photoshop 自带的“云彩”滤镜与扭曲滤镜制作的水的波纹特效。



图 T2.10 通道调板



图 T2.11 滤镜的使用——水的波纹特效

3. 制作倒影

(1) 打开两幅小狗图像，如图 T2.12 所示。我们把第二幅图中的小狗添加到第一幅图中，让它们做个伴。由于是在水边，所以在制作时要考虑给第二只小狗制作水中倒影。

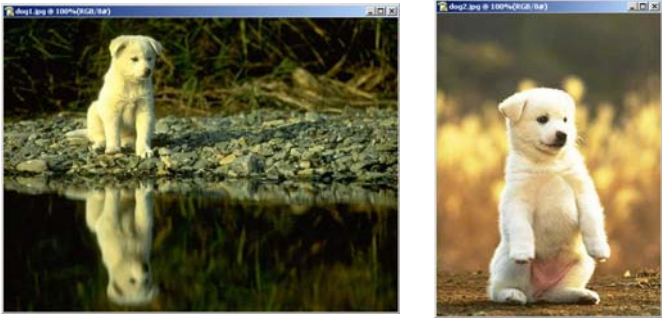


图 T2.12 两幅小狗图像

(2) 切换到第二幅图，使用磁性套索工具选中小狗，如图 T2.13 所示，然后按 Ctrl+C 组合键，将选区中的图像复制到剪切板。

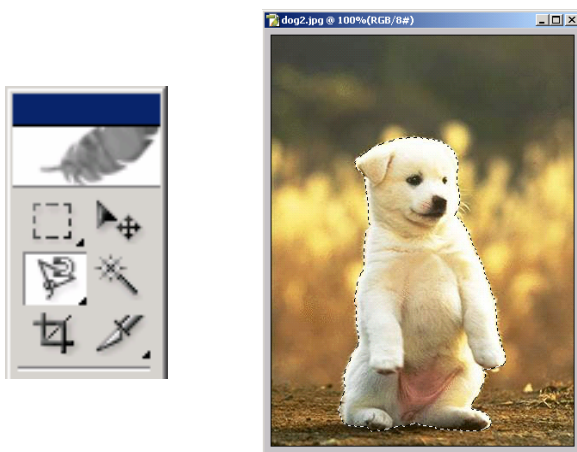


图 T2.13 用磁性套索工具选中小狗

(3) 切换到第一幅图像，按 **Ctrl+V** 组合键，把在第二幅图像中选中的小狗复制到新图层中。将新图层命名为 **dog2**，如图 T2.14 所示。

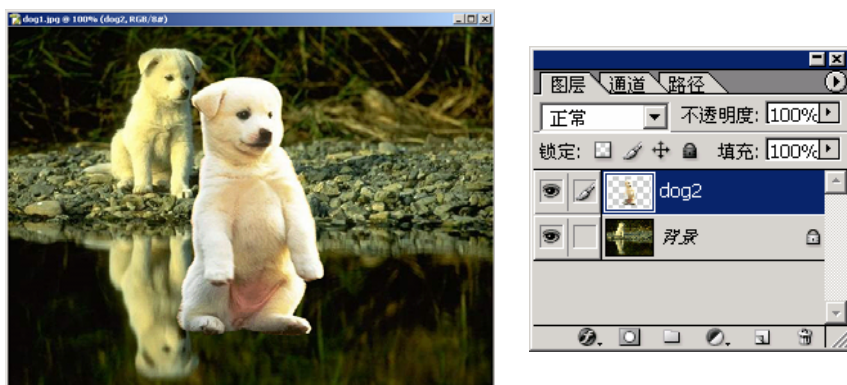


图 T2.14 复制选中的小狗到新图层

(4) 选择菜单项“编辑”→“变换”→“缩放”，对 **dog2** 层进行缩放变换，调整复制小狗的大小，在工具属性栏上单击“√”确认，如图 T2.15 所示。

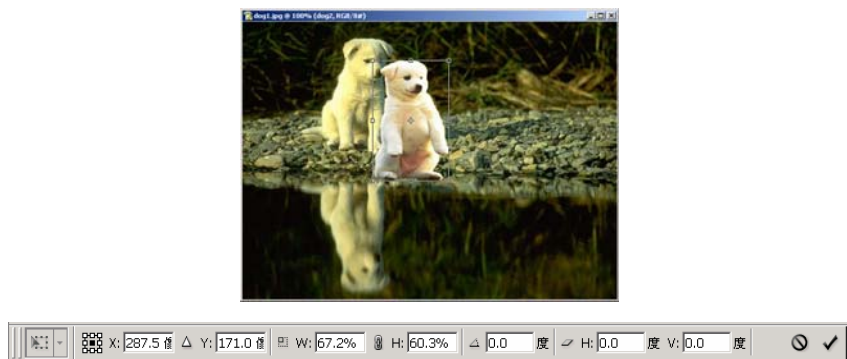


图 T2.15 调整小狗尺寸

(5) 使用移动工具调整 dog2 的位置，使之和第一只小狗并排站在水边，如图 T2.16 所示。



图 T2.16 使用移动工具调整 dog2 的位置

(6) 将 dog2 层拖动到图层面板上的“创建新的图层”按钮，复制 dog2 层。将复制出来的图层命名为 reflection，如图 T2.17 所示。

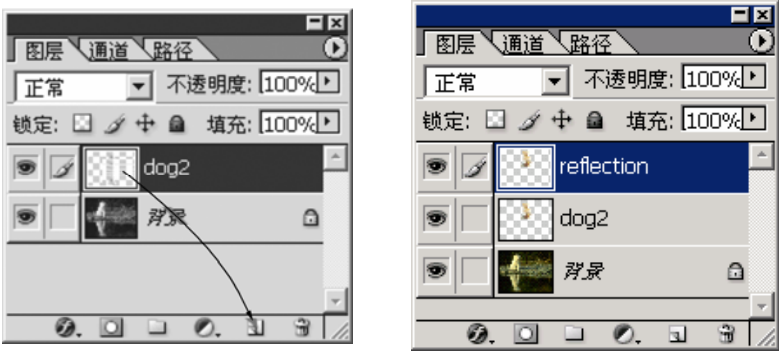


图 T2.17 复制图层

(7) 确认当前层为 reflection 层，选择菜单项“编辑”→“变换”→“垂直翻转”，对该层图像进行垂直翻转。

(8) 使用移动工具移动翻转后的 reflection 层，使它处于 dog2 的倒影位置，如图 T2.18 所示。

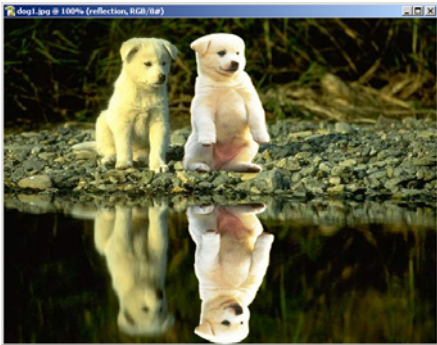


图 T2.18 制作小狗的倒影镜像

(9) 选择菜单项“滤镜”→“模糊”→“高斯模糊”，对 reflection 层进行高斯模糊，模

糊半径为 2 像素，如图 T2.19 所示。

(10) 此时倒影效果已制作完毕，如图 T2.20 所示。

倒影的制作主要用到了图层的功能。倒影其实是原图像的一个副本，只考虑到它们之间的映像关系，所以对它进行了垂直翻转。另外，通常倒影一般要比原图像模糊些，故使用了模糊滤镜对它进行模糊处理。



图 T2.19 高斯模糊滤镜



图 T2.20 倒影效果图

4. 制作蓝天绿草

(1) 选择菜单“文件” → “新建”，新建一个文档，如图 T2.21 所示。

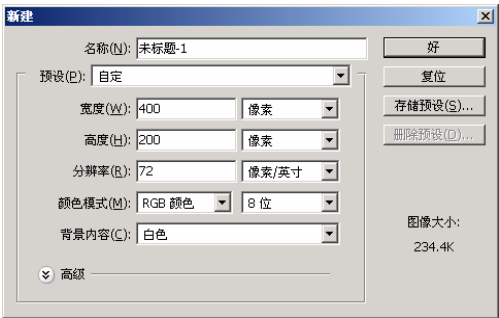


图 T2.21 新建文档

(2) 分别单击工具箱上的前景色、背景色设置按钮，设置前景色为白色，背景色 R，G，B 的值分别为 0，174，255，如图 T2.22 所示。



图 T2.22 前景色、背景色设置

(3) 选择菜单“滤镜”→“渲染”→“云彩”，效果如图 T2.23 所示。



图 T2.23 云彩效果

(4) 在工具箱上选择画笔工具，并在工具属性栏中选择草型画笔，如图 T2.24 所示。

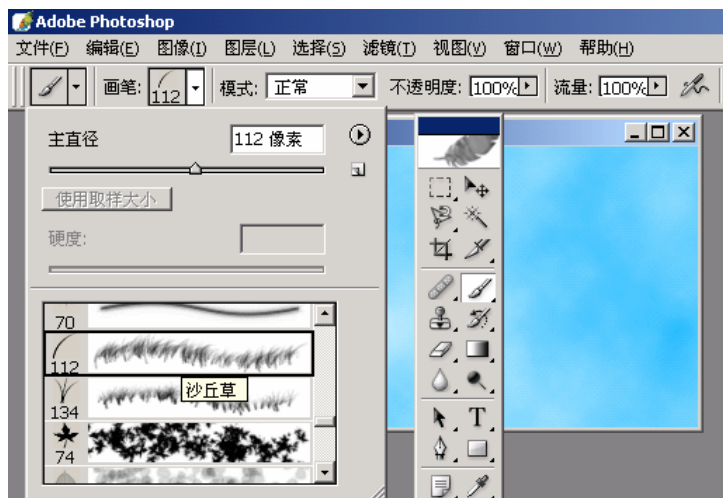


图 T2.24 选择草型画笔

(5) 重新设置前景色 (R, G, B 的值分别为 7, 105, 2)、背景色 (R, G, B 的值分别为 10, 176, 2), 如图 T2.25 所示。



图 T2.25 前景色 (左)、背景色 (右) 设置

(6) 用画笔工具在画布上刷, 最终效果如图 T2.26 所示。



图 T2.26 蓝天绿草效果图

5. 制作邮票

(1) 选择菜单“文件”→“打开”，打开一张图像，如图 T2.27 所示。



图 T2.27 打开一张图像

(2) 为了更好地显示效果, 单击工具箱上的背景色按钮, 将背景颜色设为橘红色 (R, G, B 的值分别为 165, 69, 24), 如图 T2.28 所示。

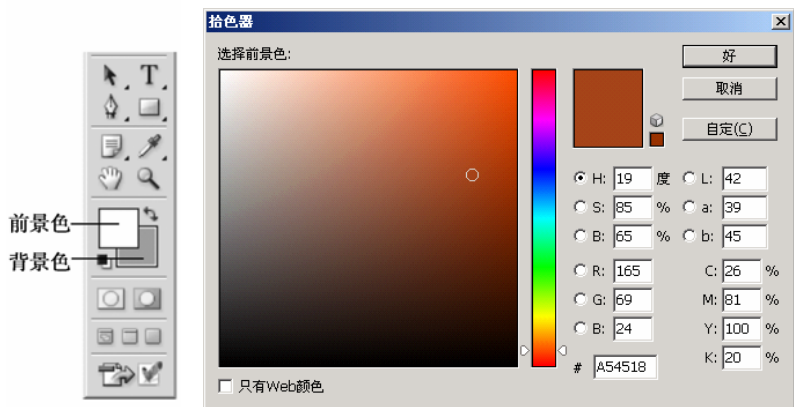


图 T2.28 设置背景色

(3) 按 **Ctrl+A** 组合键全选画布。

(4) 选择菜单“图层”→“新建”→“通过剪切的图层”，将选取区剪切到图层 1，如图 T2.29 所示。

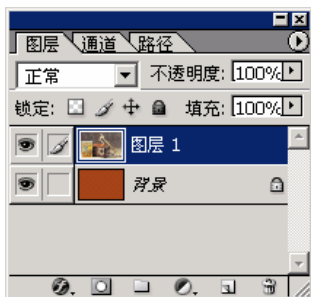


图 T2.29 新建图层

(5) 在工具箱中选择矩形选框工具，在图像窗口建立略小于画布的选区，如图 T2.30 所示。



图 T2.30 矩形选区

(6) 选择菜单“选择”→“反选”，将选区反选。选择菜单“编辑”→“填充”，在出现的对话框中设置用白色填充，单击“好”按钮，将选区用白色填充，如图 T2.31 所示。



图 T2.31 用白色填充选区

(7) 按 **Ctrl+D** 组合键将选区取消。按 **Ctrl** 键，在图层面板中单击图层 1，将图层 1 不透明区域作为选区载入。

(8) 在图层面板中单击“路径”标签，打开路径面板。在路径面板中单击“从选区生成工作路径”按钮（第 4 个按钮），通过选区生成工作路径，如图 T2.32 所示。

(9) 按下工具箱上的“橡皮擦”工具，单击窗口右上角的画笔，在“画笔笔尖形状”对话框中选择实心圆点笔尖形状，设置直径为 10 像素，间距为 150%。如图 T2.33 所示。



图 T2.32 通过选区生成工作路径

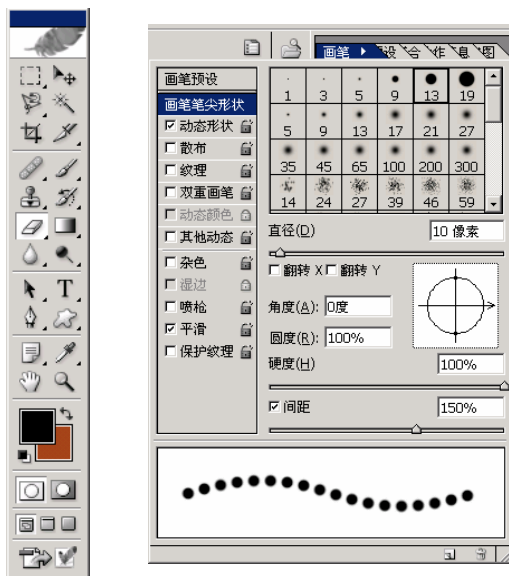


图 T2.33 画笔笔尖形状设置

(10) 在路径面板中单击“用画笔描边路径”按钮（第 2 个按钮），用画笔描边，效果如图 T2.34 所示。



图 T2.34 用画笔描边

(11) 在路径面板中单击“删除当前路径”按钮（最后一个按钮），删除当前路径，如图 T2.35 所示。

(12) 单击“图层”标签，回到图层面板。选择菜单“图像”→“画布大小”，在出现的“画布大小”对话框中设置宽度的百分比为 110%，高度的百分比为 110%，定位于中间。单击“好”按钮，将画布扩大，如图 T2.36 所示。



图 T2.35 删除当前路径



图 T2.36 “画布大小”对话框

(13) 单击图层 1 使其为当前层，选择菜单“图层”→“图层样式”→“投影”，出现“图层样式”对话框，设置距离为 15 像素，大小为 8 像素，其余的为默认值，单击“好”按钮，如图 T2.37 所示，效果如图 T2.38 所示。

(14) 在工具箱中选择文字工具，在工具属性栏上设置文字字体为宋体，大小为 30，颜色为黑色，消除锯齿方式为浑厚，其余的为默认值，如图 T2.39 所示，然后在适当位置输入文字“80 分”，在工具属性栏上单击“√”确认，最终效果如图 T2.40 所示。

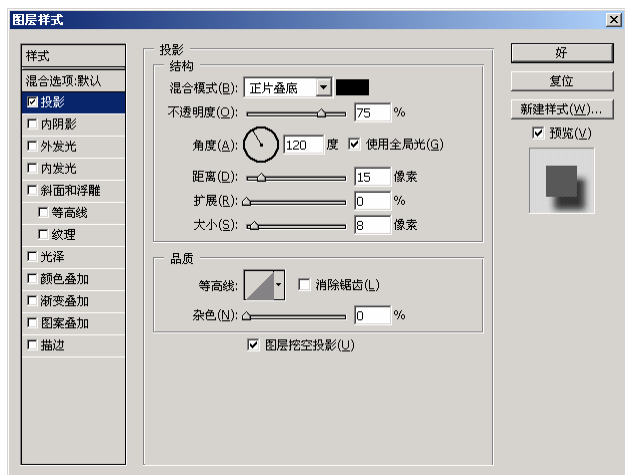


图 T2.37 “图层样式” 对话框



图 T2.38 投影效果



图 T2.39 文字工具属性栏设置



图 T2.40 邮票效果图

Photoshop 功能强大,我们通过几个简单实例介绍了其基本操作,包括图像的编辑和处理,图层、路径及滤镜的使用等。Photoshop 擅长于利用基本图像素材进行再创作,得到精美的设计作品。

实验思考

- (1) 数字图像的获取方法主要有几种? 分别是什么?
- (2) 图像文件的格式会影响数据量, 为什么?
- (3) 决定数字图像质量的技术参数有哪些?
- (4) Photoshop 软件的主要功能是什么? 基本操作有哪些?
- (5) 如何对图像进行复制、删除、剪裁、变形?
- (6) 用画笔工具能绘制枫叶吗?
- (7) 什么是图层? 其作用是什么?
- (8) 什么是图层样式? 都可以添加哪些效果?
- (9) 什么是滤镜? 其作用是什么?

实验3 数字视频

实验目的

- (1) 了解视频处理软件的相关知识，掌握视频处理的基本方法；
- (2) 熟悉 Premiere Pro 软件的主要功能及使用；
- (3) 掌握简单的加工处理视频文件的方法。

实验内容

T3.1 视频的拍摄与Premiere的使用

Premiere 是 Adobe 公司的专业非线性编辑软件。Premiere 提供与线性编辑机一致的操作方式，可以组接多种格式的视频和图像，提供多种镜头切换方式、视频叠加方式，可对图像的色调和亮度等色彩参数进行调整，方便地在视频图像上添加字幕或徽标，也可以进行音频的编辑和合成，很方便地为图像配音或为语音添加音乐配音，支持多种格式的视频输出。

在本实验中，我们将在若干视频文件中截取一些想要的片段，拼接起来，加上伴音和字幕，最后输出成想要的格式从而完成制作。

1. 导入实验所需制作的素材

启动后选择 New Project，然后在出现的对话框中选择 DV-PAL 中的 Standard 48kHz，在对话框底部选择保存工程的路径，并为工程取一个名字如 MyVideo 后单击“确定”按钮，得到如图 T3.1 所示的界面（Adobe Premiere Pro）。

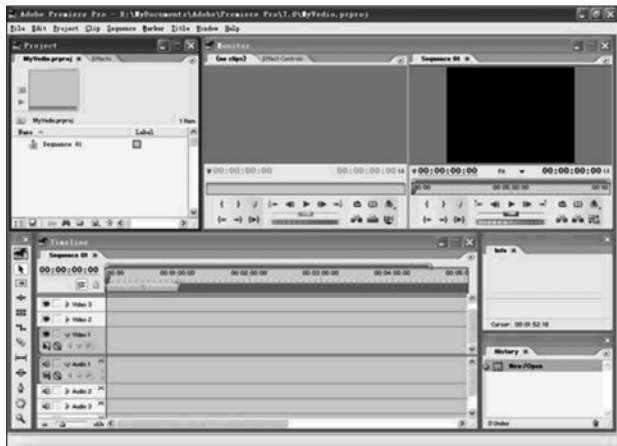


图 T3.1 Adobe Premiere Pro 的初始界面

下面计划导入两段影像和一幅图片，并分别从两段影像中各截取一个片段与导入的图片拼合成一段视频。要导入需要加工的素材，在“File”菜单中“选择 Import…”，然后选择需要导入的文件即可。注意，Premiere 支持多种视频、音频和图像文件格式，但不支持 Rm 或 Rm vb 格式。成功导入后，在“Project”窗口中可以看到素材的列表，如图 T3.2 所示。

当用鼠标单击某一项素材的时候，在“Project”窗口的上部会显示出该素材的简单摘要，如图 T3.2 中显示了 AVSEQ04.mpg 文件是一个电影文件，画面大小为 352×288、每秒 25 帧、包含 44.1kHz、16 位的立体声，并且可以在小窗口中浏览播放视频。

也可以从数码摄像机中得到视频/音频（数码摄像机的操作参见具体的说明书）。录制完毕后，需要导入到计算机中时，用 IEEE 1394 线将数码摄像机和计算机连接（计算机上需要有 IEEE 1394 接口）。将数码摄像机置于播放状态。启动 Premiere，在“File”菜单中选择“Capture”。此时出现捕捉窗口，通过窗口中的按钮便可以控制摄像机的播放操作。单击捕捉窗口中的“录制”按钮，便可以将摄像机 DV 带上的影像以 AVI 格式存储到计算机上（注意，该 AVI 是未经压缩的格式，每 10 分钟的影像会占用 2.5GB 左右的空间，需要准备足够的硬盘空间）。



图 T3.2 “Project”窗口中素材的列表

2. 组织并裁剪视频

对视频进行编辑需要将视频放置到“Timeline”窗口中。如果整段视频都需使用，则简单地将该视频文件直接从“Project”窗口拖动到“Timeline”窗口即可。如果只使用素材中的一段，一个方法是双击该视频，则它会出现现在“Monitor”窗口中，如图 T3.3 所示。



图 T3.3 “Monitor” 窗口

可以在需要使用片段的开始处单击“设置入点”按钮，在开始处做一个标记。接下来拖动播放滑块或者使用播放按钮找到片段的结尾处，再单击“设置出点”设置片段结束标记。随后单击“插入”按钮将片段插入到“Timeline”窗口，如图 T3.4 所示。

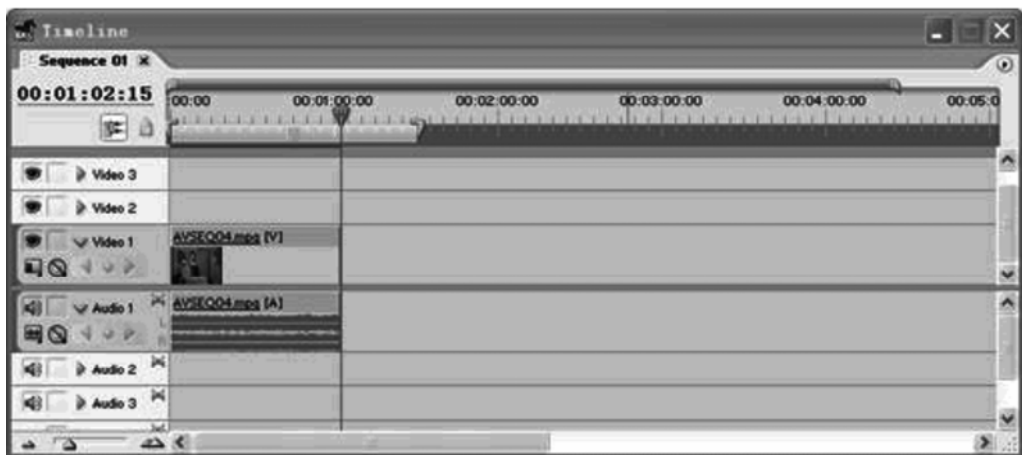


图 T3.4 “Timeline” 窗口，编辑工作主要在这个窗口完成

可以看到，在窗口中默认有 3 个 Video 轨道和 3 个 Audio 轨道。片段插入到哪一条轨道取决于哪一条是“当前”轨道。要使某条轨道成为“当前”轨道，只需在轨道的头部用鼠标单击即可。在最终的输出中，上方轨道中的内容将盖住下方的内容。

在窗口的上部是时间标尺，时间标尺的滑块上有一根红色的时间线。插入点即从该时间线开始插入。可以拖动滑块改变时间线的位置。重复上面的步骤，在结尾处插入第二个片段。随后将图片由“Project”窗口拖动到 Video2 轨道上。

一段视频结束，另一端视频紧接着开始，这就是所谓的镜头切换。为了使镜头切换衔接自然或更加有趣，可以使用各种过渡效果（Transitions）。

要运用过渡效果，可选择“Project”窗口中的“Effects”标签页，则出现特效面板，在特效面板中选择“Video Transition”文件夹，打开后可以看到详细分类的文件夹，单击任意一个扩展标志，则会显示一组过渡效果。选择合适的过渡效果后，拖到时间轴上两段视频的

重叠处，Premiere 会自动确定过渡长度以匹配过渡部分。

3. 加入字幕

可以在影片中加入自己的字幕。在“File”菜单中选择“New”，在弹出的子菜单中选择“Title”，会出现如图 T3.5 所示的字幕编辑窗口。在窗口左边的工具栏中选择文字工具（和 Photoshop 中的图标是一样的），在画面上单击，便可以输入文字。注意后面的画面仅是相对位置的参考，并不意味着字幕将出现在这幅画面上。

在字幕编辑窗口右边的 Object 栏目中可以选择字体、颜色等设置，请自行实验。在左边的工具栏中还有画图形的工具，选择一两种工具，试验一下它们的用途。

完成了字幕之后关闭字幕编辑窗口，关闭时会提示你将字幕保存为一个字幕文件（以 .pttl 文件名结尾）。选择适当的文件名和保存路径，保存字幕文件。

完成了保存工作之后字幕文件自动加入工程，此时应该可以在“Project”窗口看到刚才保存的字幕文件。将该字幕文件拖动到“Timeline”窗口的 Video3 轨道上。

选中 Video3 轨道上的字幕，当鼠标移动到表示字幕条块的边缘的时候，鼠标形状会改变成“[”或者“]”式样。此时可以改变字幕持续时间的长短，而字幕条块开始处所对应的时间点就是字幕开始出现的时间。

这样设计的字幕在屏幕上静止的。要使字幕滚动，参考图 T3.5，在字幕编辑窗口的左上角有选项：Title Type，选择不同的类型可以产生不同的滚动效果，请自行完成。



图 T3.5 字幕编辑窗口

4. 伴音的处理

如果打算采用原来视频所带有的伴音，则可以跳过此节。如果想为电影配上新的声音，则首先需要删除旧的伴音，也就是需要将 Audio 轨道上的内容删除。在删除之前，需要断开

Audio 和 Video 之间的连接。选中需要操作的视频和音频，在上面单击右键，从弹出的菜单中选择“Unlink Audio and Video”。做完这一步的操作之后，会发现此时可以单独选择音频或视频了。选择要删除的音频，按下 Delete 键。如果需要删除多个片段的音频，需要对每一个片段重复上面的操作。

如果已经有录制好的音频文件，如在 Windows 录音机中录制的 Wav 文件，可以利用“File”菜单中的“Import”命令导入到“Project”窗口，再将该音频文件拖入到 Audio 轨道上，调整好起始位置即可。伴音也可以通过 Premiere 录制，方法如下：

- (1) 在 Windows 中设置好麦克风，确保麦克风工作正常。
- (2) 在 Premiere 的“Window”菜单中选择“Audio Mixer”，出现 Premiere 的“Audio Mixer”窗口，如图 T3.6 所示。
- (3) 可以将录制的音频放置到任何一个 Audio 轨道上，在要放置的轨道上单击“麦克风”按钮，本例放置在了轨道 1 上。
- (4) 在“Timeline”窗口拖动时间滑块，放置开始播放录音的位置。
- (5) 单击“录制”按钮，此时，录制按钮开始闪动，准备开始录音。
- (6) 单击“播放”按钮，开始录音。录音完成后单击“停止”按钮，录音结束。

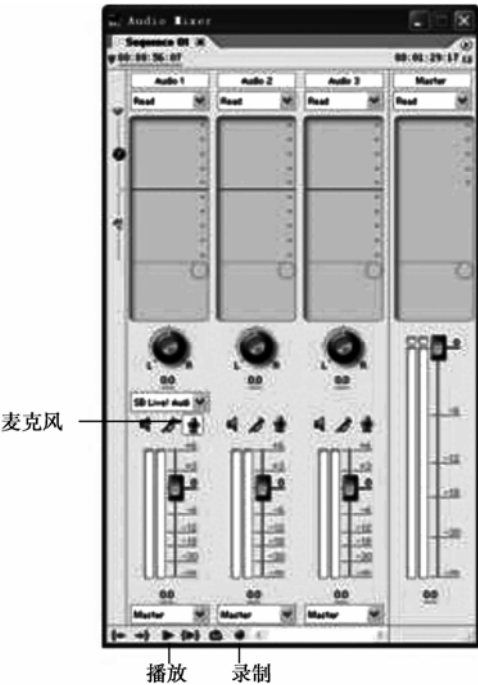


图 T3.6 Premiere 的混音器 (Audio Mixer)

可以看到一个新的音频文件被加入到了“Project”窗口中。同时在“Timeline”窗口中也加入了一个新的音频片段。此时的“Timeline”窗口如图 T3.7 所示。

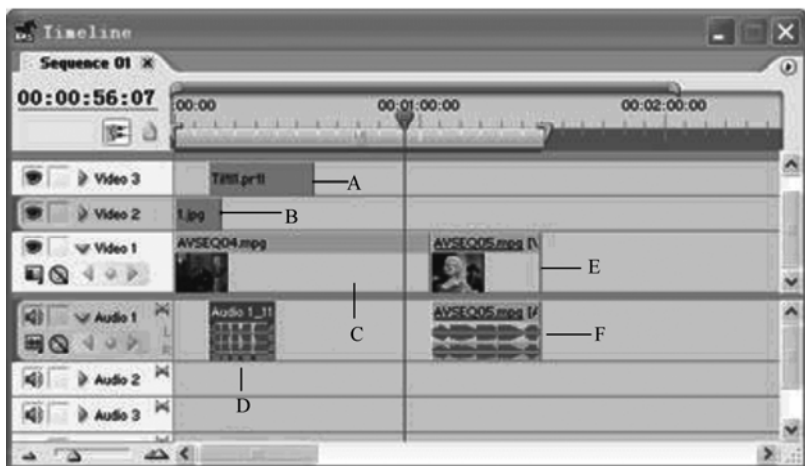


图 T3.7 制作后的“Timeline”窗口

A 是字幕文件；B 是一幅静态图像；C 是第一个片段的视频部分，音频已被删除；D 是通过麦克风录制的音频；E 是第 2 个片段的视频；F 是第 2 个片段的原有音频。

5. 输出电影

现在可以将最终的成果输出成一部完整的电影了。在输出前可以拖动时间滑块在 Monitor 窗口中预览最终的 movie，满意后再输出。Premiere 支持多种输出格式，可以输出成 mpeg 格式，也可以输出为 Windows 的 Avi 格式，还可以输出为 rm 或者 rmvb 格式。在这里，将它输出为 Windows 的 avi 格式，其余的格式请自行实验。

在“File”菜单中选择“Export”，再选择“Movie...”（注意：rm 和 rmvb 格式在“Export”菜单下的“Adobe Media Encoder”中）。在弹出的保存对话框中单击 Setting 按钮，弹出如图 T3.8 所示的输出选项对话框。

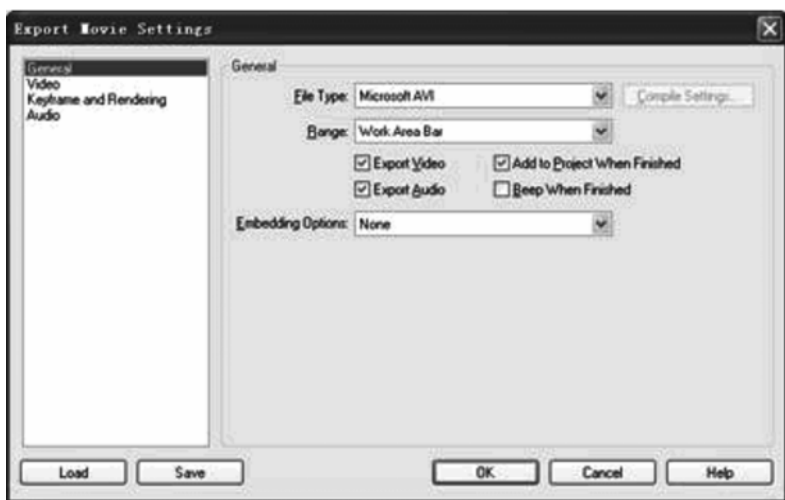


图 T3.8 输出选项对话框

在“File Type”中选择“Microsoft AVI”，在“Range”中选择“Work Area Bar”。接下来

在对话框的左边选择第二项“Video”，对话框出现相应的 Video 选项，如图 T3.9 所示。

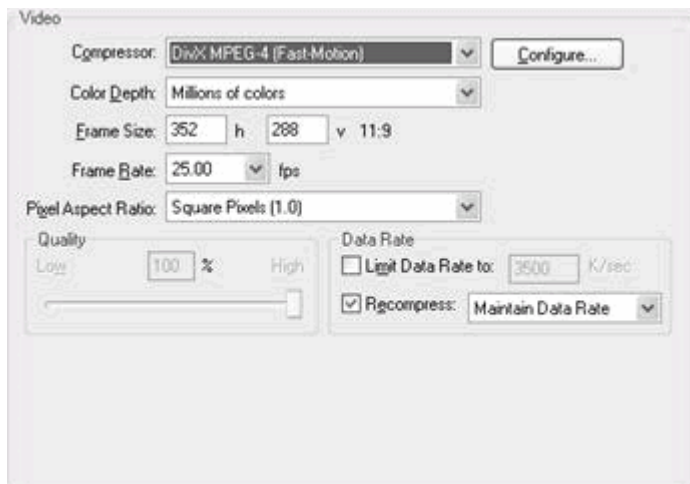


图 T3.9 Video 选项

在“Compressor”中选择一个编码/压缩器（取决于你的机器是否安装了该编码器），在“Frame Size”中设置输出画面的大小，并在“Frame Rate”中设置帧率。

其余的设置使用默认设置，单击“OK”按钮关闭配置对话框回到保存对话框。选择文件的存储路径和文件名后单击保存按钮。此时 Premiere 开始输出最终的文件，这也许需要较长的时间。Premiere 将给出估算的剩余时间（建议此时不要让计算机再运行其他程序）。

实验思考

- (1) Premiere 软件的主要功能是什么？常用窗口有哪些？其主要作用是什么？
- (2) 如何向项目中添加素材？
- (3) 如何在项目中实现滚动字幕？
- (4) 如何在项目中加入声音文件？
- (5) 如何保存和输出电影文件？
- (6) 如何实现素材之间的转换？尝试其他滤镜效果，过渡效果，透明效果等。

实验 4 数 字 动 画

实验目的

- (1) 了解 GIF 动画的特色及基本原理;
- (2) 熟悉 Ulead Gif Animator 软件的主要功能;
- (3) 了解 Flash 动画的特色及基本原理;
- (4) 熟悉 Flash 动画制作软件的主要功能。

实验内容

T4.1 Ulead GIF Animator制作GIF动画实例

Ulead Gif Animator 是一个简单、快速、灵活,功能强大的 GIF 动画编辑软件,同时,也是网页设计辅助工具,还可以作为 Photoshop 的插件使用。丰富而强大的内制动画选项,让更方便地制作符合要求的 GIF 动画。

Ulead GIF Animator 5.05 功能很强大,使用它可做出真彩色环境下的 GIF 动画,得到色彩斑斓的动画。功能更强的工具和更为灵活的工作环境,也使我们的动画制作愈显轻松。其界面如图 T4.1 所示。



图 T4.1 Ulead GIF Animator 5.05 界面

1. 初识Ulead GIF

(1) 菜单

菜单栏包括文件、编辑、查看、对象、帧、过滤器、视频 F/X 和帮助，8 个菜单分别完成从图像创建导入到图层滤镜和视频特殊效果的功能。

(2) 标准工具栏

Ulead GIF Animator 的工具栏上包含标准的 Windows 应用程序按钮，即新建、打开、保存、撤销、恢复、剪切、复制、粘贴按钮，使用方法和其他 Windows 应用程序一样。在工具栏上还有一些 Ulead GIF Animator 专用的按钮：添加图像（Add Images）、添加视频（Add Video）等。

(3) 工具面板

工具面板提供了一整套画图的基本工具，可以进行简单的画图和图像修改。

(4) 对象设置面板

对象设置面板能够对作品出现的对象进行属性的修改。

(5) 帧控制面板

帧控制面板管理动画中出现的每一帧，支持帧的添加、删除、设置帧之间的间隔时间等操作。帧控制按钮包括帧的播放、停止、跳转等控制。

(6) 工作区

工作区中包括“编辑”、“优化”、“预览”面板。

① 编辑：在这个区域中可以查看 GIF 动画的结构，还可以针对每一幅 GIF 动画图像进行编辑。

② 优化：在设计动画过程中，当帧数较多时，这个动画文件可能会较大，不适合在网络上传播，所以 GIF Animator 中提供了对动画进行优化的功能。这种功能的原理是将帧的图像分割成若干块，在帧之间变换时，相同的块将在第二个帧中被删除，达到减小文件大小的目的。所以我们要对自己创建的动画再进行一下优化，在 GIF Animator 中预设了 7 个方案，可以将图像优化成色彩数不同的 GIF 动画，相应地优化后的质量也是不同的。同时还可以对优化的各种参数进行设置，在优化后程序还会报告优化的效果。

单击“Optimize”按钮，进入优化设置窗口，如图 T4.2 所示。在这个窗口的左侧是原始图像，右侧是系统按照默认优化设置所优化的图像。另外有一个窗口是 Color Palette，如图 T4.3 所示。此窗口中显示了优化后图像中使用的颜色。如果对优化的效果不满意，或者对优化后所取得的大小不满意，可以进行有关像素的设置。

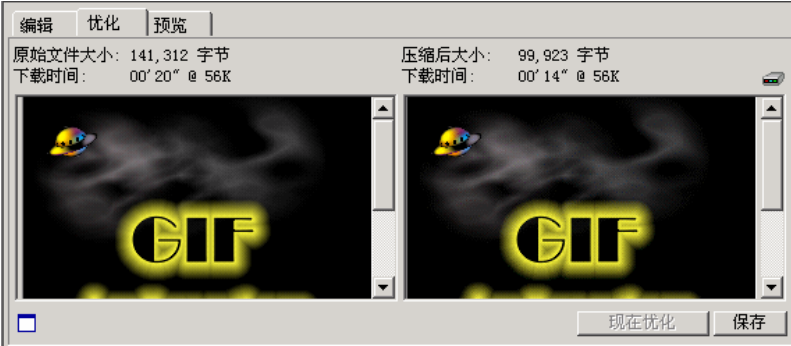


图 T4.2 Ulead GIF Animator 优化设置窗口

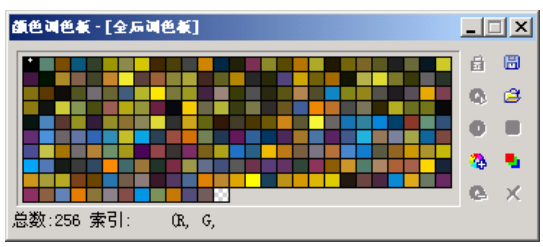


图 T4.3 Ulead GIF Animator Color Palette 窗口

③ 预览：完成一切的制作后，预览动画的效果。

(7) 渲染效果

Ulead GIF Animator 不仅可以支持最简单的动画，即把所有的图片夹杂在一起，还可以在图片中加入特殊的效果，如三维效果等。所有的渲染效果均可在 Vido F/X 菜单栏中找到。不同的渲染效果见表 T4.1。

表 T4.1 Ulead GIF Animator 的渲染效果

效 果		功 能
3D	Gate-3D（三维立体）	在动画中实现两个帧之间的动画三维过渡效果，可以在对话框中设置过渡方向、填充颜色和柔化过渡边界的参数
Build	Diagonal-Build（倾斜过渡）	按照不同的倾斜对角线路径，逐渐地按照不规则块的方式实现两个帧之间的过渡
Clock	Sweep Clock（时钟式过渡）	按照顺时针和逆时针方式实现过渡效果
F/X	Diamond F/X（菱形 F/X 动画）	菱形效果
	Iris F/X（彩虹 F/X 动画）	十字心形的彩虹效果
	Mosaic F/X（马赛克 F/X 动画）	马赛克元素效果
	Power off F/X（关机 F/X 动画）	显示器关机效果
Film	Flap B-Film（分割的倾斜翻页效果）	将原来的帧分为 4 份，分别从中心开始倾斜地实现与目标帧的过渡
	Progressive Film（前进式翻页效果）	也是将原来的帧分为 4 份，分别从中心开始以前进式的方式实现到目标帧的过渡
	Turn Page-Film（整页翻页效果）	以整页翻页的效果来实现两个帧之间的过渡，同时可以设置翻页的方向等参数
Peel	Turn Page-Peel（整页翻页的剥落效果）	以剥落的效果实现两个帧之间的过渡，与上一个翻页效果略有不同
Push	Run and Stop Push（移动到停止间的推动效果）	为了消除两个帧之间生硬的动画元素移动，使用这种效果可以生成好看的动态效果
Roll	Side Roll（整面翻转效果）	整面翻转效果
Slide	Bar-Slide（条状滑动效果）	可以从不同的方向用两个条状的滑动实现过渡
Stretch	Cross Zoom-Stretch（交叉缩放拉伸效果）	交叉缩放拉伸效果
Wipe	Star Wipe（星形擦除效果）	星形擦除效果
2D Mapping	Crop…（修剪效果）	修剪效果

效 果		功 能
Camera Len	Gradient（倾斜效果）	在原帧中加入模糊的镜头，可以设定移动方向等参数
	Mirror（镜像效果）	对原帧中的明显的对象加以处理，实现两个帧之间的过渡
	Zoom Motion（缩放动作效果）	缩放效果
Darkroom	Hue&Saturation（色调饱和）	设置图像中的颜色的变化效果
Nature Painting	Charcoal（木炭效果）	木炭效果
Special	Add Noise（加入噪声）	在原帧中加入噪声干扰效果
	Wind（风）	风的效果

当动画制作完成后，Ulead GIF Animator 最直接的输出就是把动画保存为 GIF，可以让它在网页中随意地被插入。Ulead GIF Animator 同样支持其他文件格式，如 PSD 及非常流行的 Flash 格式（SWF）。这些文件格式只能用于浏览，如果再次进行编辑则需要重新来过。因此，Ulead GIF Animator 自己开发了一种保留 GIF 动画信息的文件格式：UGA，当需要再次修改这个动态 GIF 时就可以调用这个文件，然后再进行修改。

Ulead GIF Animator 同时还支持输出为网页格式和输出为桌面项目。网页格式在这里不用向大家解释，桌面项目是一个很有意思的功能，它利用了 Windows 中自带的桌面自定义功能，把动态的 GIF 动画放置在桌面上。Ulead GIF Animator 还可以直接输出为 EXE 文件，并且能够添加消息框、声音文件和文字信息，让用户只需双击此文件，即可观看到动态图片的内容。

2. 制作跳动的青蛙（逐帧动画）

首先使用图像处理软件生成 5 幅相同尺寸的青蛙跳跃过程的图像文件，如图 T4.4 所示。



图 T4.4 一组 GIF 格式的图像素材文件



- （1）打开图像。单击标准工具栏上的 （打开图像）按钮，在弹出的对话框中选择“frog1.gif”，打开第一幅青蛙图像。
- （2）添加图像。单击标准工具栏上的 （添加图像）按钮，在“添加图像”对话框中，按住 Ctrl 键单击选择多幅图像 frog2.gif, frog3.gif, frog4.gif, frog5.gif，如图 T4.5 所示。
- （3）添加帧。单击动画控制条中的“添加帧”按钮，如图 T4.6 所示。
- （4）在对象设置面板上单击“显示/隐藏”按钮（外观有一只眼睛），选择需要的第 2 幅青蛙动作图像，如图 T4.7 所示。



图 T4.5 “添加图像”对话框



图 T4.6 添加帧

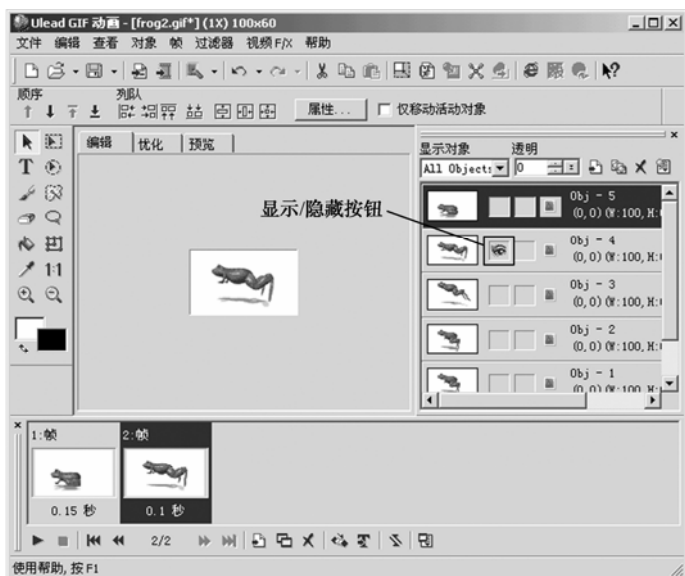


图 T4.7 选择需要的图像

(5) 重复动作 (4) 插入青蛙跳跃的所有帧，效果如图 T4.8 所示。

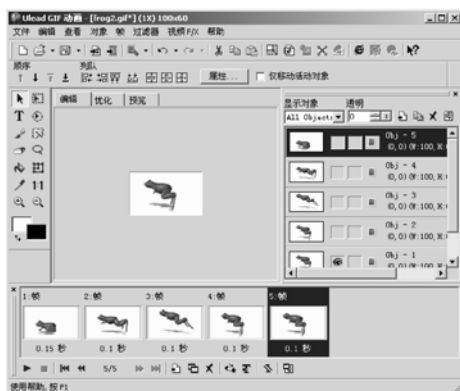


图 T4.8 插入所有帧

(6) 查看效果。单击工作区上的“预览”按钮，查看动画效果，如图 T4.9 所示。



图 T4.9 预览效果

(7) 调整效果。单击工作区的“编辑”按钮，回到编辑状态，GIF Animator 的动画制作面板中默认第一帧的时间是 0.1s 切换一帧，如果达不到希望的效果，要对延迟时间进行调整。可双击一帧，在“画面帧属性”对话框中，调整延迟时间，如图 T4.10 所示。



(a) 帧的延迟时间



(b) 画面帧属性设置

图 T4.10 调整帧之间的延迟时间

(8) 保存动画文件。选择菜单“文件”→“另存为”→“GIF 文件”，将动画保存为 GIF 文件格式，如图 T4.11 所示。

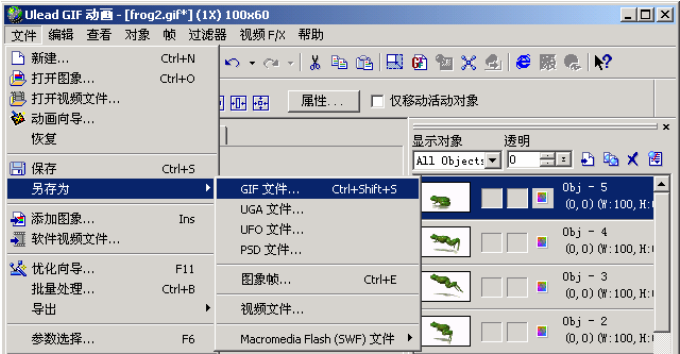


图 T4.11 保存 GIF 动画

3. 制作“屏幕移动效果”(关键帧动画)

(1) 新建文件。选择菜单“文件”→“新建”，设定文档大小为 200 像素×140 像素（小于要设置移动效果的图片）。

(2) 添加图像。单击标准工具栏上的（添加图像）按钮，添加要制作移动效果的图片。

(3) 添加帧。单击动画控制条中的“相同帧”按钮，在工作区编辑面板中稍稍移动一下图的位置，如图 T4.12 所示。



图 T4.12 添加相同帧

(4) 重复步骤(3)两次。

(5) 选中第 1 帧，按住 Shift 键，在第 4 帧上单击（4 帧全选中），右键单击“画面帧属性”菜单，在“画面帧属性”对话框中将延迟时间改为 30，如图 T4.13 所示。

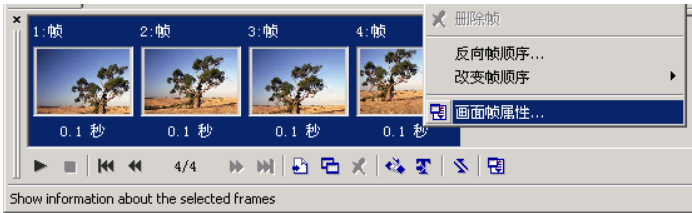


图 T4.13 设置帧延迟时间

(6) 单击第 1 帧，单击动画控制条中的“之间”按钮，按如图 T4.14 所示进行设置，开始帧是 1 表示第 1 帧，结束帧是 2 表示第 2 帧，意思是说在第 1 帧和第 2 帧之间插入 5 帧。



图 T4.14 帧“之间”设置

(7) 单击第 7 帧，单击“之间”按钮，开始帧改为 7，结束帧改为 8。

(8) 单击第 13 帧，单击“之间”按钮，开始帧改为 13，结束帧改为 14。

(9) 将第 1 至第 19 帧全部选中，单击动画控制条中的“相同帧”按钮，复制前面的帧，然后右键单击，选择“反向帧顺序”菜单，在弹出的对话框中选择“选定帧相反顺序”，如图 T4.15 所示，翻转复制帧的前后顺序，使得整个动画连贯。



图 T4.15 翻转帧的前后顺序

(10) 保存动画文件。选择菜单“文件”→“另存为”→“GIF 文件”，将动画保存为 GIF 文件格式。

4. 制作“图形变化效果”

(1) 选择菜单“文件”→“打开图像”，打开一幅图像。

(2) 选择菜单“视频 F/X”→“F/X”→“Iris-F/X”，打开“添加效果”对话框，添加星形擦除效果（在菜单“视频 F/X”下有很多动画效果，可以组合成自己喜欢的形式），如图 T4.16 所示。

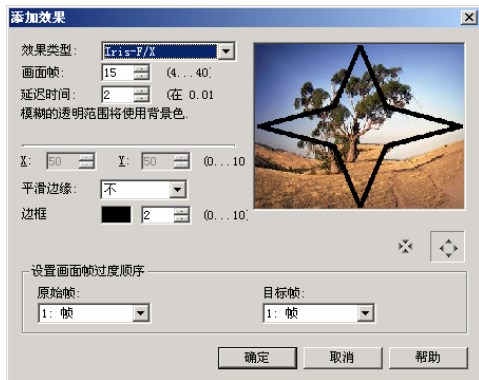


图 T4.16 “添加效果”对话框

(3) 保存动画文件。选择菜单“文件”→“另存为”→“GIF 文件”，将动画保存为 GIF 文件格式。

5. 制作“文字效果”

(1) 选择菜单“文件”→“打开图像”，打开一幅图像。

(2) 单击动画控制条中的“添加文本条”按钮，如图 T4.17 所示。



图 T4.17 添加文本条

(3) 在出现的对话框中选中“文本”标签，输入汉字并选定自己喜欢的字体、颜色、大小，如图 T4.18 所示。



图 T4.18 文本属性设置

(4) 选中“效果”标签，“进入场景”选择“垂直合并”，画面帧设为 10（即用 10 帧表现进入场景的动画）；“退出场景”选择“放大”，画面帧设为 15，如图 T4.19 所示。

(5) 选中“画面帧控制”标签，延迟时间设为 10，在“分配到画面帧”前打上钩，如图 T4.20 所示。

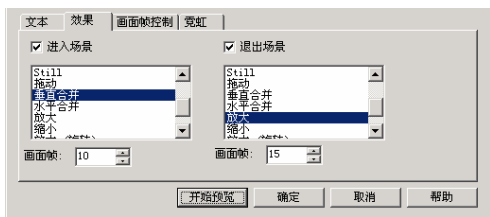


图 T4.19 文字效果设置



图 T4.20 帧控制设置

(6) 单击“确定”按钮，选择“创建为文本条”。

(7) 保存动画文件。选择菜单“文件”→“另存为”→“GIF 文件”，将动画保存为 GIF 文件格式。

T4.2 Flash制作动画实例

Flash MX 2004 是矢量图形编辑和动画创作软件，它通过符号、按钮、层、帧、场景等

一系列组合，能够让用户集成图形、声音、动画、影像文件等各种多媒体素材，像用砖头造房子一样制作出形式简洁、内容丰富、交互性强的极富感染力的动画作品，给用户提供了一个方便、广阔的动画制作平台，其界面如图 T4.21 所示。

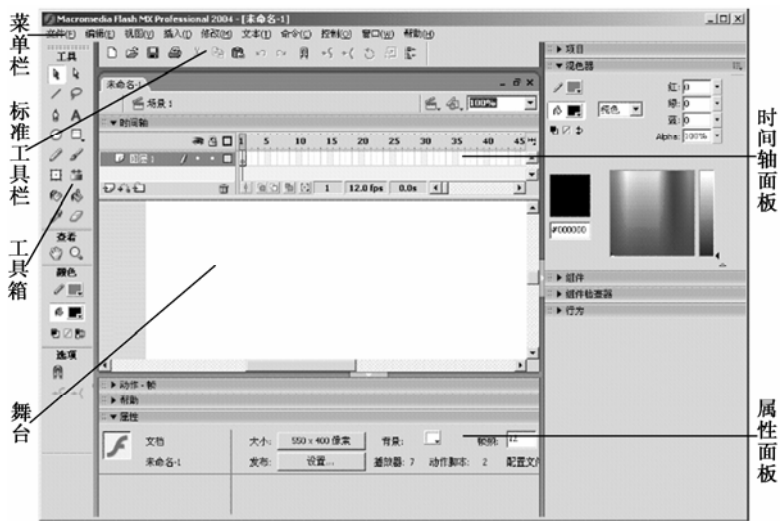


图 T4.21 Flash MX 界面



图 T4.22 Flash MX 工具箱

1. 初识Flash

(1) 菜单栏

菜单栏包括除绘图命令以外的绝大多数 Flash 命令。

(2) 标准工具栏

标准工具栏不仅包括文件和编辑菜单中的常用命令，还包括 Flash 中的一些工具按钮，如对象的旋转、缩放等。选择“窗口”→“工具栏”→“主工具栏”命令可显示或隐藏标准工具栏。

(3) 工具箱

Flash MX 工具箱 (如图 T4.22 所示) 包括用于创建、放置和修改文本与图形的工具。它位于窗口的左侧，可以使用鼠标将其拖至窗口的任意位置。工具箱由工具、查看、颜色和选项 4 个区域组成，其中的选项区域用于显示工具所包含的功能键选项，当用户选择不同的工具时，选项区域中就会出现与之相对应的功能键。可分别选择线条、矩形等工具，在舞台中绘制简单图形，验证其功能。

(4) 时间轴面板

时间轴用于组织和控制影片内容在一定时间内播放的层数和帧数。“时间轴”面板位于标准工具栏下方，如图 T4.23 所示。选择“窗口”→“时间轴”命令，可

打开或关闭“时间轴”面板。

时间轴的各组成部分如下。

① 时间轴的主要组件是图层、帧。与胶片一样，Flash 影片也将时长分为帧。图层就像层叠在一起的幻灯胶片一样，每个图层都包含一个显示在舞台中的不同图像。

② 文档中的图层列在时间轴左侧的列中。每个图层中包含的帧显示在该图层名右侧的一行中。时间轴顶部的时间轴标题显示帧编号。

③ 时间轴状态显示在时间轴的底部，它指示所选的帧编号、当前帧频，以及到当前帧为止的运行时间。

④ 在帧视图弹出菜单中可以更改帧的显示方式，也可以在时间轴中显示帧内容的缩略图。时间轴可以显示影片中哪些地方有动画，包括逐帧动画、补间动画和运动路径。

⑤ 时间轴的图层部分中的控件可以隐藏或显示、锁定或解锁图层，以及将图层内容显示为轮廓。

⑥ 可以在时间轴中插入、删除、选择和移动帧，也可以将帧拖到同一图层中的不同位置，或拖到不同的图层中。

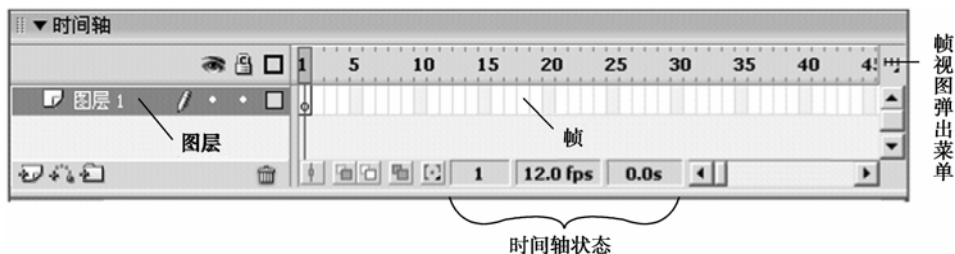


图 T4.23 “时间轴”面板

(5) “属性”面板

“属性”面板是 Flash MX 新增的面板，它集成了 Flash 5 浮动面板中的常用选项。当在工作区中选取某一对象或在绘图工具栏中选择某些工具时，属性面板中将显示与它们对应的属性。例如，单击绘图工具栏中的“文本工具”，屏幕下方即可显示如图 T4.24 所示的文本工具“属性”面板。



图 T4.24 文本工具“属性”面板

(6) 舞台

舞台是创作影片中各个帧的内容的区域，用户可以在其中自由地绘图，也可以在其中安排导入的插图，编辑和显示动画，并配合控制工具栏的按钮演示动画。

2. 制作“逐笔写字”（逐帧动画）

逐笔写字动画描述的是一种书写的效果，运行效果如图 T4.25 所示。

徐徐徐徐徐徐徐徐

图 T4.25 逐笔写字动画的运行效果


- (1) 选择菜单“文件”→“新建”命令，选择“Flash 文档”新建一个文件。
- (2) 在时间轴中的图层 1 的第 1 帧处创建起始关键帧。选择文本工具  在舞台上单击，在文本框中输入“徐”，在属性面板中设置字体为“文鼎中特广告体”，80 号字，黑色，如图 T4.26 所示。



图 T4.26 设置文本属性



- (3) 分离文字。用箭头工具  选中文字块，选择菜单“修改”→“分离”命令，将文字变成图像。
- (4) 创建其他关键帧。在当前帧的后面一帧右键单击选择“插入关键帧”，在场景中出现和前面一帧一样的汉字，如图 T4.27 所示。
- (5) 选择橡皮擦工具 ，在工具箱面板的橡皮擦选项中，设置橡皮擦的大小，擦除当前帧汉字的最后一笔画，如图 T4.28 所示。
- (6) 重复步骤 (4)、(5) 直到最后一帧汉字的笔画全部擦完。



图 T4.27 插入关键帧



图 T4.28 擦除一次后的汉字“徐”

- (7) 交换关键帧。上述步骤操作完毕后，效果是倒序书写的效果，即先写“徐”的最后一笔。为了达到正常的书写效果，在时间轴上第 1 帧处按住鼠标左键不放，拖动到最后一帧，选中所有的帧（如图 T4.29 所示），选择菜单“修改”→“时间轴”→“翻转帧”命令，将所有关键帧前后交换。
- (8) 选择菜单“控制”→“测试影片”命令，测试该动画，产生一种逐笔画写字的效果。
- (9) 选择菜单“文件”→“保存”命令，保存 Flash 动画源文件 (*.fla)。选择菜单“文件”→“发布”命令，将导出动画文件 (*.swf) 和网页文件 (*.html)。

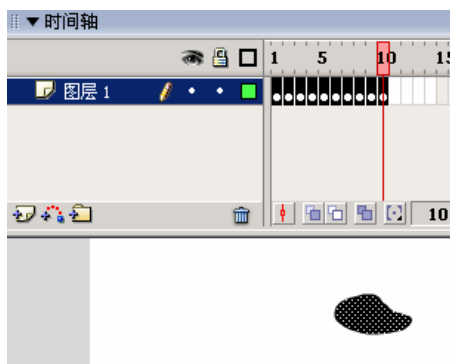


图 T4.29 选中所有的帧

3. 制作“滴水效果”

该动画描述水滴掉下来后引起小小波浪的效果，如图 T4.30 所示。

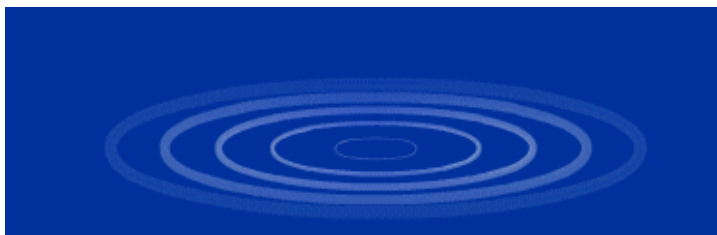


图 T4.30 滴水效果动画运行效果

(1) 选择“文件”→“新建”命令，选择“Flash 文档”新建一个文件。设置文档属性窗口背景色为蓝色，如图 T4.31 所示。

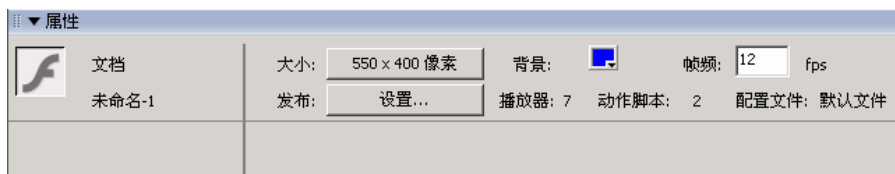


图 T4.31 文档属性设置

(2) 创建水滴元件。选择菜单“插入”→“新建元件”命令，弹出“创建新元件”对话框，为新元件设置名称为“水滴”，行为选“图形”，如图 T4.32 所示。

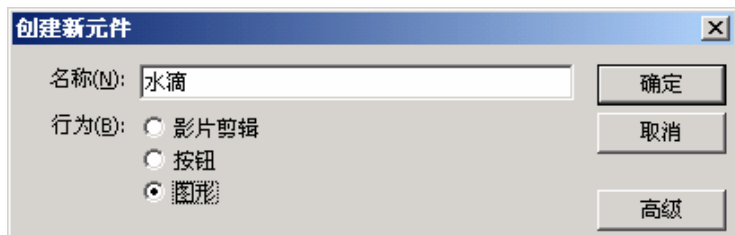



图 T4.32 “创建新元件”对话框

选择椭圆工具，在工具箱面板的颜色选项中，设置椭圆的轮廓颜色为白色，无填充色，在编辑区中央绘制一个圆形，如图 T4.33 所示。

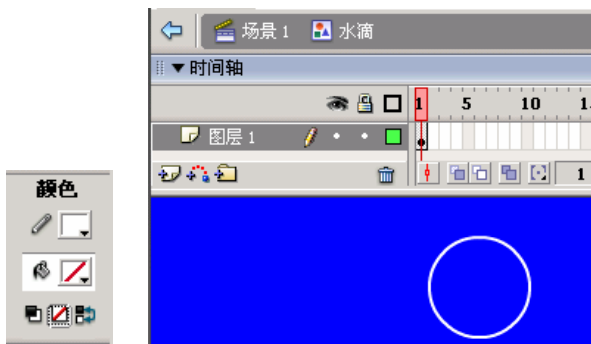


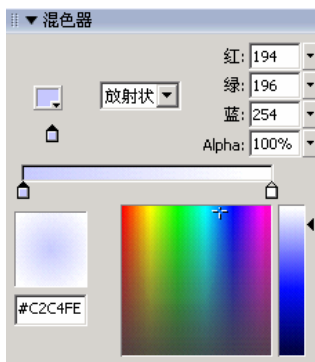


图 T4.33 绘制圆

选择颜料桶工具，在右边的面板中选择混色器面板，调制填充颜色，选择“放射状”填充方式，在渐变色上设置两个小颜料桶，左边小颜料桶颜色十六进制值为#C2C4FE，右边小颜料桶颜色设置为白色，如图 T4.34（a）所示，用颜料桶对绘制的圆形进行填充。选择工具对绘制的圆形进行变形（当选择工具靠近圆形时箭头下面会出现半圆标记，即可进行调整），使之变成水滴的形状，如图 T4.34（b）所示。




（a）混色器设置



（b）用选择工具调整的水滴图

图 T4.34 混色器和调整水滴图

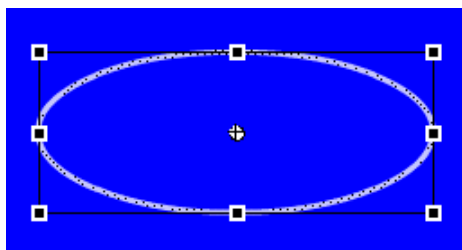
（3）创建波浪动画元件。该元件主要表达一个水波从小慢慢变大的效果。同样选择菜单“插入”→“新建元件”命令，为新元件设置名称为“波浪动画”，行为选“影片剪辑”。进入该元件的编辑模式。选择椭圆工具（去掉填充色的颜色，轮廓色和“水滴”元件中的填充色一样）在编辑区绘制一个椭圆，如图 T4.35（a）所示。

然后在第 30 帧处右键单击，选择“插入关键帧”命令。在第 30 帧的编辑区中，用任意变形工具放大椭圆，注意要保持椭圆的中心不变（拖动鼠标的同时按下 Alt+Shift 组合键），如图 T4.35（b）所示。

最后，单击第一帧，打开该帧的“属性”面板，在补间下拉列表框中选择“形状”选项后，生成一条黑色的实线箭头，该层的背景色变为浅绿色，表示动画创建成功，可以移动播放头来预览该元件的动画效果，如图 T4.36 所示。



(a) 绘制小波浪



(b) 绘制大波浪

图 T4.35 创建波浪动画元件

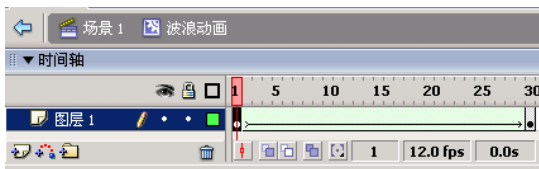
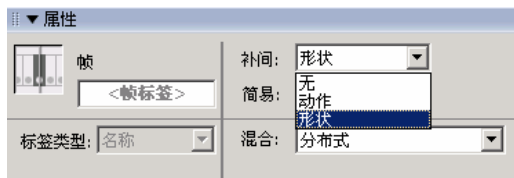


图 T4.36 创建波浪动画

(4) 制作水滴落下效果。单击 **场景 1** 切换到场景 1，从右边“库”面板（若未打开，可选择菜单“窗口”→“库”命令）中将水滴元件拖到编辑区上部分，如图 T4.37 所示。

在第 7 帧处右击后选择“插入关键帧”命令，将编辑区中的水滴元件向下垂直移动，表示水往下滴的效果，如图 T4.38 所示。

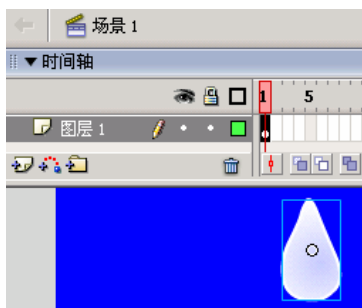


图 T4.37 水滴在上面

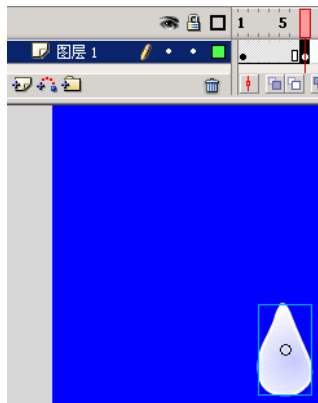


图 T4.38 水滴在下面

最后，单击第一帧，打开该帧的“属性”面板，在补间下拉列表框中选择“动作”选项后，生成一条黑色的实线箭头，该层的背景色变为浅蓝色，表示动画创建成功，可以移动播放头来预览该元件的动画效果，如图 T4.39 所示。

(5) 制作层层波浪。右键单击图层 1，在弹出的菜单中选择“插入图层”命令，新建一图层 2，在第 7 帧处插入关键帧，从“库”面板中拖入波浪动画元件，调整其位置如图 T4.40 所示。



图 T4.39 水滴落下效果

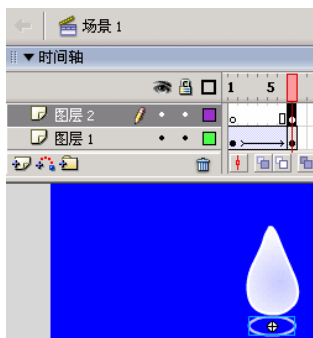


图 T4.40 设置波浪位置

在图层 2 第 36 帧处插入一关键帧，用选择工具选中波浪动画元件，打开其“属性”面板，设置颜色选项为 Alpha，值为 0%，如图 T4.41 所示。

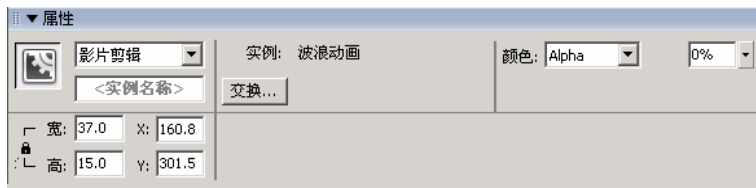


图 T4.41 设置元件的 Alpha 值

然后，单击图层 2 第 1 帧，打开该帧的“属性”面板，在补间下拉列表框中选择“动作”选项后，第 1 层波浪动画就制作完成了。

依次类推可添加图层 3，4，5，6，再制作 4 层波浪，每层波浪依次往后推移 5 帧即可，但要注意保持各层的波浪动画元件的中心相同。设置好的图层如图 T4.42 所示。

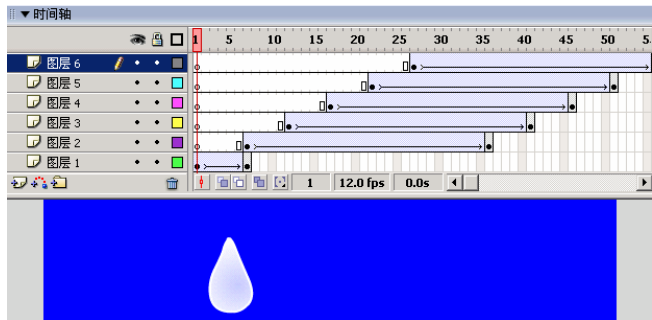


图 T4.42 设置好的滴水效果的所有图层

最后测试影片，保存文件，该动画就制作完毕了。

4. 制作“大雪纷飞效果”

该动画描述大雪纷飞的效果，如图 T4.43 所示。



图 T4.43 大雪纷飞动画运行效果

(1) 选择“文件”→“新建”命令，选择“Flash 文档”新建一个文件。单击“属性”面板中的“大小”按钮，打开“文档属性”对话框，设置场景大小为 500px×240px，背景为 #00105A，帧频为 18fps，如图 T4.44 所示。

(2) 按 Ctrl+F8 组合键打开“创建新元件”对话框，新建一个名为“雪片”的图形元件，单击“椭圆”工具，在“雪片”元件的编辑场景中按 Shift 键拉出一个直径为 5 像素的圆，再用“箭头”工具稍微调整一下，画一个雪片的形状，如图 T4.45 所示。



图 T4.44 文档属性设置

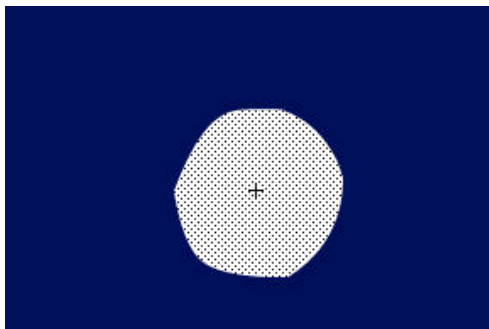


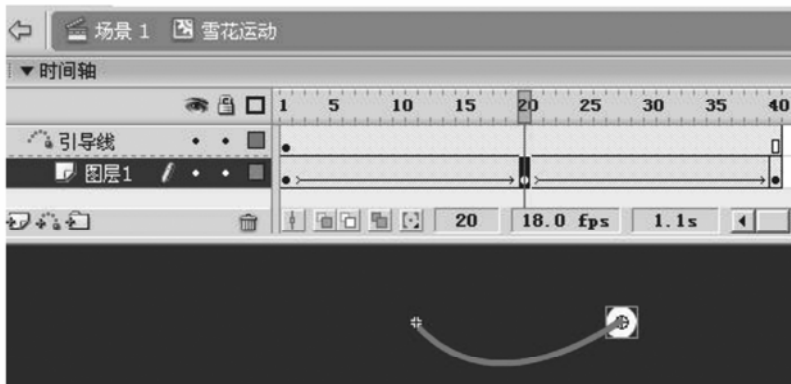
图 T4.45 雪片元件

(3) 新建一个名为“雪花运动”的影片剪辑元件，将“库”面板中的“雪片”元件拖入该层的场景中。选中“雪片”层的第 40 帧按下 F6 键插入关键帧，接着右键单击第 1 帧选择“创建补间动画”命令，再选中第 20 帧插入一个关键帧。

(4) 右键单击图层 1，在弹出的菜单中选择“添加引导层”命令为图层 1 建立一个引导层。在引导层里用铅笔工具画一条曲线，将第 1 帧和第 40 帧的“雪片”元件吸附在曲线的左端，将第 20 帧的“雪片”元件吸附在曲线的右端。图 T4.46 (a)、图 T4.46 (b)、图 T4.46 (c) 分别演示了在第 1 帧、20 帧、40 帧雪片元件的位置（如果不好调整可以用缩放工具放大显示比例）。



(a) 第1帧雪片元件的位置




(b) 第20帧雪片元件的位置



(c) 第40帧雪片元件的位置

图 T4.46 雪片元件的位置

(5) 单击  切换到场景 1，把“库”中的“雪花运动”元件拖入到场景中，选中第 3 帧按下 F5 键插入帧。选中场景中的“雪花运动”元件，在“属性”面板里给影片剪辑“雪花影片”取个实例名“snow”，如图 T4.47 所示。

(6) 在场景 1 中新建一个图层，双击图层名称改名为“action”，分别选中层中的第 1，2，3 帧按 F7 键插入空白关键帧，这时你也可以把图层 1 改名为“snow”。按 F9 键打开“动作”面板，分别选中“action”图层的第 1，2，3 帧，加入脚本，如图 T4.48 所示。



图 T4.47 影片剪辑属性设置

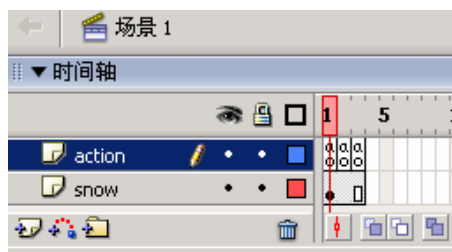


图 T4.48 为关键帧加入脚本

① 第 1 帧中的代码如下：

```
//变量 flakes 用来控制雪花的数量，即密度
//snow 为雪片影片剪辑的实例名，_visible 属性隐藏母本
n = 1;
flakes = 100;
snow._visible = 0;
```

② 第 2 帧中的代码如下：

```
i = 1;
while (i <= flakes)
{
    //控制雪花飘落速度 (this 对象的作用是引用影片剪辑实例)
    this["snow" add i]._y += Math.random () * 2 + 2;
    i++;
}
if (n == flakes)
{
    n = 1;
} else
{
    // Math.random () 方法取 0~1 之间的随机数
    ssize = Math.random () * 50 + 30;
    //复制雪花，并为新的雪花取名为"snow" add n
    duplicateMovieClip ("snow", "snow" add n, n);
    //控制雪花飘落范围 (500 为场景的宽度)
    this["snow" add n]._x = Math.random () * 500;
    //雪花随机变换大小
    this["snow" add n]._xscale = ssize;
    this["snow" add n]._yscale = ssize;
    n++;
}
}
```

③ 第 3 帧中的代码如下：

```
//返回到第 2 帧播放
gotoAndPlay (2);
```

效果到这里就完成了，本实例主要介绍了 Flash 中脚本的运用，自从 Flash 5 推出 ActionScript 之后，动画编程受到越来越多的重视。要使动画复杂度进一步提高，制作出优秀的作品还需要一定的编程基础。

实验思考

- (1) 什么是 GIF 动画？只支持 256 色吗？
- (2) Ulead GIF Animator 软件的主要特点及功能是什么？基本操作有哪些？文件格式有哪些？
- (3) Ulead GIF Animator 软件自带的图形变化效果有哪些？文字效果有哪些？
- (4) Flash 软件的主要特点及功能是什么？基本操作有哪些？文件格式有哪些？
- (5) 在 Flash 中，直接输入的文本是不能编辑形状的，怎样使它的形状也能够进行编辑？
- (6) Flash 的补间动画有哪几种类型？有什么区别？
- (7) 如何创建 Flash 的图形和影片剪辑元件？如何使用元件？
- (8) 如何对 Flash 的关键帧进行 ActionScript 脚本编程？
- (9) Flash 生成的动画是矢量动画吗？和 GIF 动画有什么不同？
- (10) 模仿大雪纷飞效果动画怎样做一个下雨动画？

实验 5 多媒体创作工具

实验目的

- (1) 了解 Authorware 软件的基本功能，熟悉其编辑环境；
- (2) 了解 Authorware 软件提供的图标按钮功能，掌握其使用方法；
- (3) 了解 Authorware 软件提供的事件响应类型，掌握交互响应的使用方法；
- (4) 掌握多媒体应用软件的基本创作手段和技巧。

实验内容

T5.1 Authorware简介

在各种多媒体创作工具中，Macromedia 公司推出的多媒体制作软件 Authorware 是不可多得的开发工具之一。它使得不具有编程能力的用户也能创作出一些高水平的多媒体作品。

Authorware 采用面向对象的设计思想，是一种基于图标（Icon）和流线（Line）的多媒体开发工具。它把众多的多媒体素材交给其他软件处理，本身则主要承担多媒体素材的集成和组织工作。

Authorware 操作简单，程序流程明了，开发效率高，并且能够结合其他多种开发工具，共同实现多媒体的功能。它易学易用，不需大量编程，使得不具有编程能力的用户也能创作出一些高水平的多媒体作品，对于非专业开发人员和专业开发人员都是一个很好的选择。

同许多 Windows 程序一样，Authorware 具有良好的用户界面，如图 T5.1 所示。

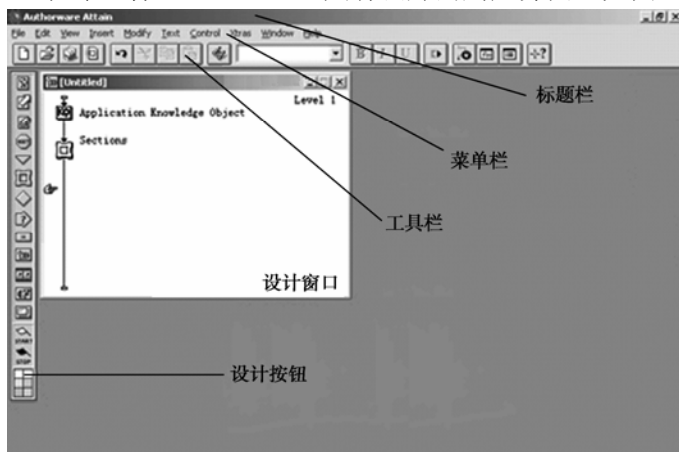


图 T5.1 Authorware 界面

(1) 工具栏

工具栏（如图 T5.2 所示）是 Authorware 窗口的组成部分，其中每个按钮实质上是菜单栏中的某一个命令，由于使用频率较高，因此被放在常用工具栏中。熟练使用常用工具栏中的按钮，可以使工作事半功倍。

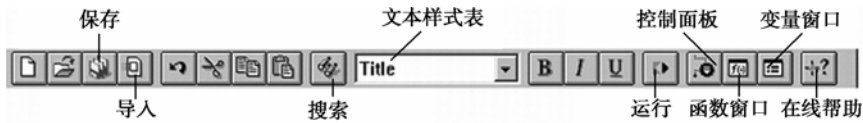


图 T5.2 常用工具栏

(2) 设计图标

设计图标在 Authorware 窗口中的左侧，如图 T5.3 所示，包括 13 个图标、开始旗、结束旗和图标调色板，是 Authorware 最特殊，也是最核心的部分。

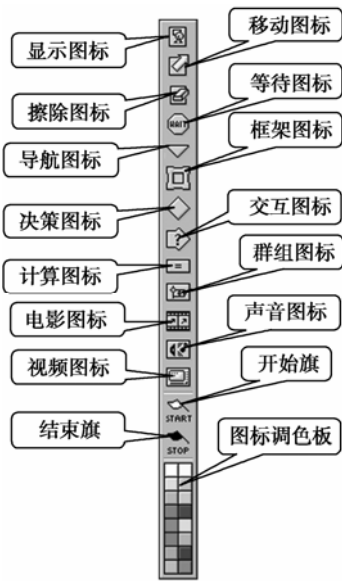












图 T5.3 设计图标

这些设计图标是 Authorware 中能够实现特定功能的功能模块（见表 T5.1），使用这些设计图标，能够完成交互式应用程序的开发。

表 T5.1 Authorware 设计图标

图 标	说 明
	显示设计图标：在显示设计图标中，用户可以在其展示窗口中输入正文或图片对象
	移位设计图标：可以移动显示对象以产生特殊的动画效果
	擦除设计图标：可以擦除显示在展示窗口中的任何对象
	等待设计图标：用于设置一段等待的时间
	导航设计图标：用于设置与任何一个附属于框架设计图标的设计图标间的一个定向链接，当程序运行到导航设计按钮时，Authorware 会自动定位到定向链接指向的页
	框架设计图标：该设计图标中的内容主要是一组导航控制图标。这组导航控制图标使用户能够设计交互式应用程序来正确访问一个指定的页

按 钮	说 明
	决策设计图标：可以用来设置一种判定逻辑结构，附属于该设计图标的其他设计图标被称为路径
	交互设计图标：用来创建一种交互作用分支结构，该分支结构是由一个交互作用设计图标和附属于该设计图标的其他设计图标所组成的
	计算设计图标：可以放置到程序的任何位置，执行数学运算和执行 Authorware 源程序
	群组设计图标：利用它可以将一组设计图标组合成一个简单的群组设计图标，每一个群组设计图标中可以包含它自己的逻辑结构
	电影设计图标：用于将一个电影文件引入到 Authorware 中，并对引入的电影文件进行一套控制来播放
	声音设计图标：用于在多媒体应用程序中引入声音文件
	视频设计图标：用于在多媒体应用程序中引入视频信息数据，然后在视频播放器上播放
	开始旗帜：用于调试用户程序，可以设置程序运行的起始点
	停止旗帜：用于调试用户程序，可以设置程序运行的终止点
	图标调色板：用来为设计图标着色，在程序的设计过程中，我们可以使用该设计图标来为流程线上的设计图标着色

(3) 设计窗口

程序设计窗口是 Authorware 的设计中心，Authorware 具有的对流程可视化编程功能，主要体现在程序设计窗口的风格上。

程序设计窗口如图 T5.4 所示，其组成如下。

标题栏：显示被编辑的程序文件名。

主流程线：一条被两个小矩形框封闭的直线，用来放置设计图标，程序执行时，沿主流程线依次执行各个设计图标。程序开始点和结束点两个小矩形，分别表示程序的开始和结束。

粘贴指针：一只小手，指示下一步设计图标在流程线上的位置。单击程序设计窗口的任意空白处，粘贴指针就会跳至相应的位置。

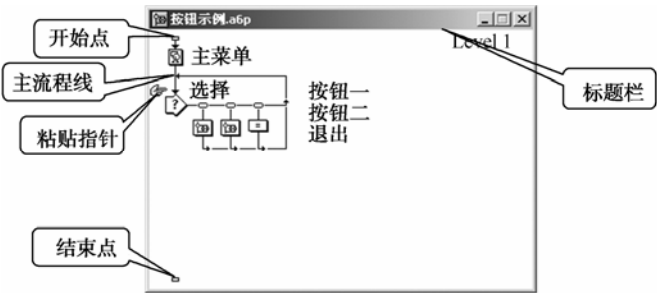


图 T5.4 设计窗口

T5.2 Authorware多媒体创作实例

1. 制作“影片介绍短片”

制作一个简单的影片介绍短片，要求加入影片画面、音乐及相关文字介绍。


(1) 从左边的设计图标栏上拖动一个显示图标到流程线上，命名为“背景”，双击该图标，弹出窗口，选择菜单栏中的“File”→“Import”，输入一张背景图，如图 T5.5 所示。



图 T5.5 输入背景图

(2) 拖动一个显示图标到流程线上，命名为“标题”，双击该图标，选择文字工具“A”，输入“我的第一个作品”，并做相关设置，如图 T5.6 所示。

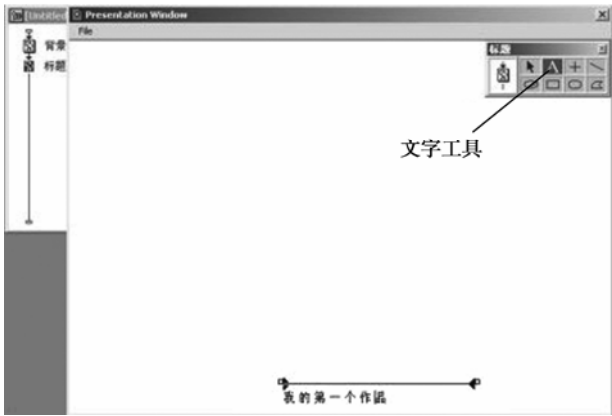


图 T5.6 输入“我的第一个作品”文字



(3) 拖动一个影片图标到流程线上，命名为“片头影片”，双击该图标，在弹出的对话框的左下角单击“Import”按钮，输入一段 MPG 文件，如图 T5.7 所示。



图 T5.7 片头影片设置

(4) 拖动一个等待图标到流程线上，命名为“等待”，双击该图标，弹出对话框设置，如图 T5.8 所示。

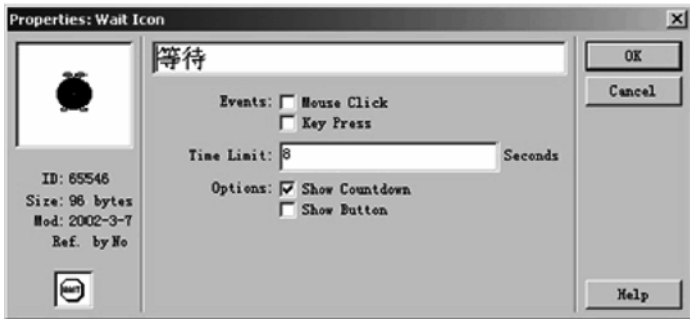


图 T5.8 等待图标设置



(5) 拖动一个擦除图标到流程线上，命名为“擦除开头画面”，双击该图标，弹出对话框设置，如图 T5.9 所示。



图 T5.9 擦除开头画面设置

单击 Transition 后的按钮进入“Erase Transition”对话框，设置擦除效果，如图 T5.10 所示。

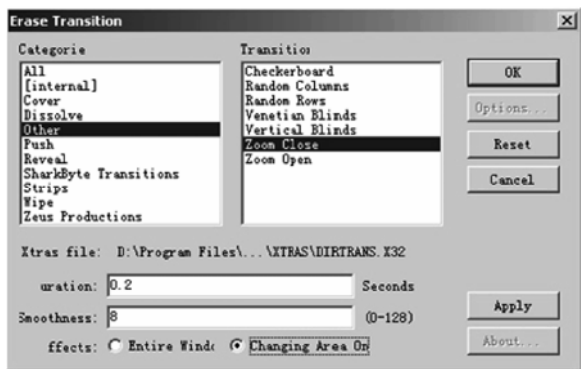


图 T5.10 “Erase Transition”对话框

(6) 拖动一个显示图标到流程线上，命名为“影片内容”。双击该图标，选择文字工具“A”，输入影片的内容介绍。


(7) 拖动一个声音图标到流程线上，命名为“介绍影片”，双击该图标，在弹出的对话框的左下角单击“Import...”按钮，输入一段 WAVE 文件，如图 T5.11 所示。



图 T5.11 影片介绍设置

(8) 完成的程序流程图如图 T5.12 所示。选择菜单“Control”→“Play”，运行程序观看效果。



图 T5.12 完成的程序流程图

从创建一个简单实例的过程可以看出，使用 Authorware 软件创作作品是非常简单直观的。Authorware 软件有代表各类多媒体信息的图标，根据软件设计的需要，将这些图标拖到流程线上，然后再往图标里加入相应内容，最后保存文件并运行，就可以看到结果。

2. 制作“拼图游戏”

本例是通过鼠标的拖曳，将 9 个小图片组成大图片，如果拖曳的目的地正确，则自动对齐排好，否则回到原位置。将一个大图分为若干块小图，可以用 Photoshop 的切片功能简单做到。

拼图游戏的运行界面如图 T5.13 所示。

拼 图 游 戏



请用鼠标将右图中的 9 个小图拖入左图中合适的位置

图 T5.13 拼图游戏的运行界面

完成的程序流程图如图 T5.14 所示。

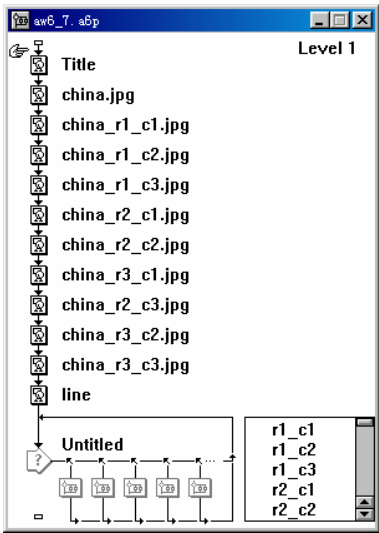


图 T5.14 完成的程序流程图

制作步骤如下：

- (1) 在流程线上拖入一个显示图标，命名为“Title”，在其合适位置输入“拼图游戏”标题和玩法介绍文字。
- (2) 选择菜单栏中的“File”→“Import”命令，或工具栏上的导入图标，单击对话框右下角的“+”，一次导入多个图像文件。将大图片和分割后的小图片共 10 个图像文件一次导入，如图 T5.15 所示。

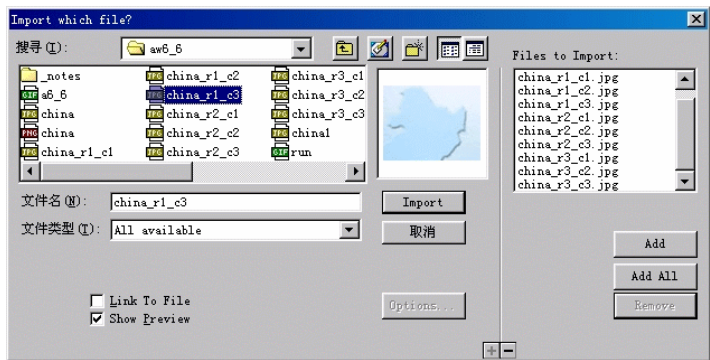


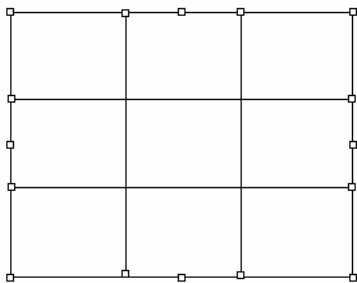


图 T5.15 导入图片

(3) 拖入一个显示图标，命名为“line”，用矩形工具和直线工具画出 9 个方块，如图 T5.16 (a) 所示，不关闭当前窗口，按住 Shift 键分别双击放有 9 个小图的显示图标，调整各小图位置，如图 T5.16 (b) 所示。



(a) 画出 9 个方块



(b) 调整小图位置

图 T5.16 小图定位

(4) 向流程线上拖入一个交互图标，命名为“Target Area”。向其右边拖入一个群组图标，命名为“r1_c1”，其作用是设计一块小图的拖动。交互响应类型选择“目标区域(Target Area)”，如图 T5.17 所示。

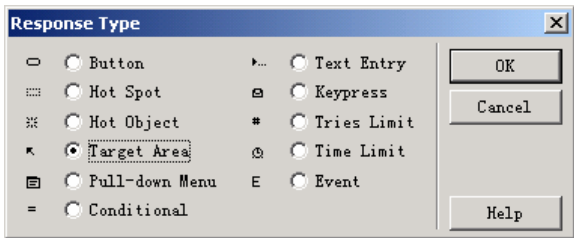

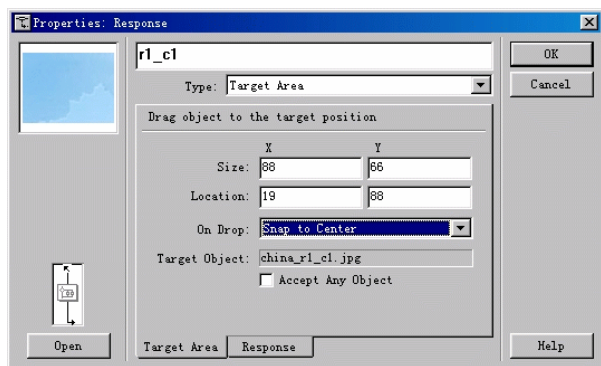


图 T5.17 选择交互响应类型

(5) 双击“china.jpg”显示图标，按住 Shift 键双击“china_r1_c1.jpg”显示图标，然后双击“r1_c1”图标上面的图标，对其响应进行详细设置。先单击图片“china_r1_c1.jpg”，然后将其拖到目的位置，调整目标区域大小，如图 T5.18 (a) 所示。在“On Drop”选项中选择“Snap to Center”（对齐目的位置的中心），如图 T5.18 (b) 所示，意为拖动图片“china_r1_c1.jpg”到目的位置松开鼠标后，图片自动对齐目的区域的中心位置。用同样方法设置其他 8 块小图片。



(a) 设定目标区域



(b) 设定响应方式

图 T5.18 响应设置

(6) 再拖一个群组图标到“Target Area”图标的右边，命名为“error”，其作用是如果小图片没有被拖到目的位置，那么还回到原来位置。对它的响应设置如图 T5.19 所示。目的区域设置为整个屏幕，不限定拖动的目标（Accept Any Object）。

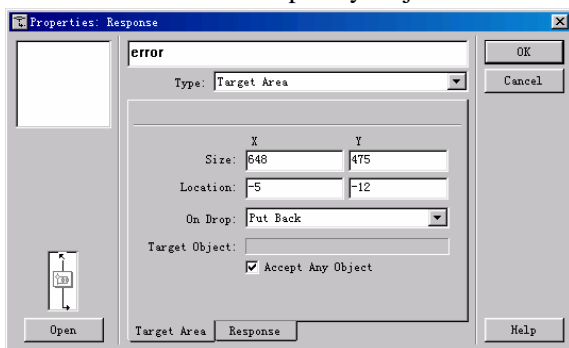


图 T5.19 响应设置

(7) 选择菜单“Control”→“Play”，运行程序观看效果。

3. 发布作品

“一键发布”（One Button Publishing）是 Authorware 6 具有的功能，这个功能可以轻松地将作品发布到 Web，CD-ROM 或局域网，使得发布 Authorware 程序非常简单。在发布之前，Authorware 将对程序中的所有图标进行扫描，找到其中用到的外部支持文件，如 Xtras，Dll 和 UCD 文件，还有 AVI，SWF 等文件，并将这些文件复制到发布后的目录中。

选择“File”→“Publish”→“Publish Settings...”命令，设置发布选项，Authorware 首先对程序中的所有图标进行扫描，如图 T5.20 所示。

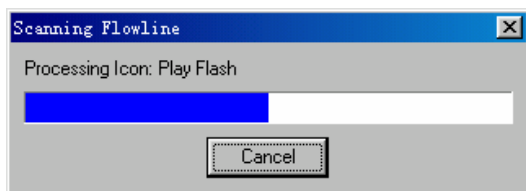


图 T5.20 扫描图标

然后出现发布设置对话框，如图 T5.21 所示。

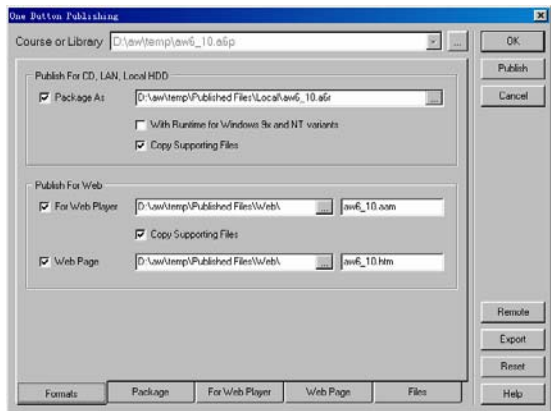


图 T5.21 发布设置对话框

“Formats” 页签是关于发布文件类型的一些设置。可以发布为带播放器的 **With Runtime** 文件（EXE 文件）、不带播放器的 **Without Runtime** 文件（A6R 文件）、使用网络播放器播放的 **For Web Player** 文件（AAM 文件）或网页（HTM 文件）文件。

“Package” 页签是关于打包文件的一些设置。如是否将库文件一同打包，是否将链接文件嵌入等。

在“**For Web Player**” 页签中，可以设置发布后每一块文件的大小，根据不同的网络连接速度，将文件分为不同大小的多个文件，使得在网速较慢时也能流畅播放。

“**Web Page**” 页签是关于发布 HTM 文件的一些选项。你的 Authorware 程序将被链接到这个 HTM 文件中，但在浏览时需要用户安装了 Authorware Web Player 才能正确浏览，如果用户机器上没有 Authorware Web Player，将提示用户下载，这一点和 Flash 的 SWF 插入网页是相似的。

在“**Files**” 页签中，可以看到当前应用程序的一些支持文件，如 Xtras，Dll 和 UCD 等，文件发布时必须将这些文件同时发布才能保证不会出现意外的错误。当然你也可以通过“**Add File (s) ...**” 按钮自己在其中手动添加一些文件，如使用说明书、帮助文档等。

以上的设置一般不需要特别设定，如果你有特殊要求，设置好的各选项还可以使用“**Export**” 保存为注册表文件（REG 文件），以方便下次使用同样的设置。

设置好后，按“**Publish**” 按钮，应用程序就成功发布了（如果不需要将程序发布给其他人，只想在本机上播放，也可以使用“**File**” → “**Publish**” → “**Package...**” 命令打包程序）。

实验思考

- (1) Authorware 软件的主要功能是什么？基本操作有哪些？
- (2) Authorware 软件的基本创作思路是什么？
- (3) 在 Authorware 软件中，如何添加文字、设置图像、设置声音及设置视频？
- (4) Authorware 软件提供的交互方式有哪些？
- (5) 如何设置响应的方式？
- (6) 如何制作一个电子相册或通讯录？

反侵权盗版声明

电子工业出版社依法对本作品享有专有出版权。任何未经权利人书面许可，复制、销售或通过信息网络传播本作品的行为；歪曲、篡改、剽窃本作品的行为，均违反《中华人民共和国著作权法》，其行为人应承担相应的民事责任和行政责任，构成犯罪的，将被依法追究刑事责任。

为了维护市场秩序，保护权利人的合法权益，我社将依法查处和打击侵权盗版的单位和个人。欢迎社会各界人士积极举报侵权盗版行为，本社将奖励举报有功人员，并保证举报人的信息不被泄露。

举报电话：(010) 88254396; (010) 88258888

传 真：(010) 88254397

E-mail: dbqq@phei.com.cn

通信地址：北京市万寿路 173 信箱

电子工业出版社总编办公室

邮 编：100036